University of California
Santa Barbara

Short-Range Millimeter-Wave Sensing and Imaging: Theory, Experiments and Super-Resolution Algorithms

A dissertation submitted in partial satisfaction
of the requirements for the degree

Doctor of Philosophy
in
Electrical and Computer Engineering

by

Babak Mamandipoor

Committee in charge:

Professor Upamanyu Madhow, Chair
Professor Amin Arbabian, Stanford University
Professor João P. Hespanha
Professor Yasamin Mostofi
Professor Kenneth Rose

September 2017

The Dissertation of Babak Mamandipoor is approved.

Professor Amin Arbabian, Stanford University

Professor João P. Hespanha

Professor Yasamin Mostofi

Professor Kenneth Rose

Professor Upamanyu Madhow, Committee Chair

August 2017

Short-Range Millimeter-Wave Sensing and Imaging: Theory, Experiments and Super-Resolution Algorithms

by

Babak Mamandipoor

To my parents, with love and gratitude.

Acknowledgements

I would like to express my sincere gratitude to Professor Upamanyu Madhow for his advice, guidance, and steady support throughout the four years of my graduate studies at UCSB. I feel extremely fortunate to have had the opportunity to work with him, and have greatly benefited from his vision, enthusiasm, and invaluable insights. I am also deeply thankful to Professor Amin Arbabian, for many insightful discussions and incisive comments that have greatly helped shape this dissertation. I would also want to thank my committee members, Professor João Hespanha, Professor Yasamin Mostofi and Professor Kenneth Rose, for their time, feedbacks and ideas for improving this work.

Many others have played a role in this research. I sincerely thank my collaborators from UCSB and Stanford; Dinesh Ramasamy, Greg Malysa, Hossein Roufarshbaf, Mahmoud Sawaby, Mojtaba Fallahpour, and Nemat Dolatsha for many valuable discussions on various research topics. I also want to thank my colleagues at WCSL lab, Anant, Andrew, Aseem, Dinesh, Faruk, Hossein, Maryam, Soorya, and Zhinus, for their company and friendship, and many interesting discussions on research and life. I am deeply thankful to my friends at UCSB Iranian graduate student association (IGSA), who brought joy and excitement to my life in Santa Barbara.

I am greatly indebted to my parents, Ozra and Mamand, for their unconditional love and support. Without their patience and encouragement, I could have never reached this far. Last, but not least, I would like to thank my brothers, Behrooz and Mehran, for their love, support, and unwavering believe in me.

Curriculum Vitæ

Babak Mamandipoor

Education

2017	Doctor of Philosophy, Electrical and Computer Engineering, University of California, Santa Barbara, USA.
2013	Master of Applied Science, Communications and Information Systems, University of Waterloo, Canada.
2010	Bachelor of Science, Electrical Engineering, Iran University of Science and Technology, Tehran, Iran.

Publications

B. Mamandipoor, A. Arbabian, U. Madhow, "Geometry-Constrained Degrees of Freedom Analysis for Imaging Systems: Monostatic and Multistatic," *to be submitted shortly.*

B. Mamandipoor, M. Fallahpour, A. Arbabian, U. Madhow, "Short-Range Millimeter Wave Imaging Using Dictionary of Spatially-Extended Objects: Theory and Experiments," *to be submitted shortly.*

(J2) **B. Mamandipoor**, D. Ramasamy, U. Madhow, "Newtonized Orthogonal Matching Pursuit: Frequency Estimation over the Continuum," *IEEE Transactions on Signal Processing (TSP)*, vol. 64, no. 19, pp. 5066-5081, Oct.1, 1 2016.

(J1) **B. Mamandipoor**, K. Moshksar, Amir K. Khandani, "Capacity-Achieving Distributions in Gaussian Multiple Access Channel With Peak Power Constraints," *IEEE Transactions on Information Theory,* vol. 60, no. 10, pp. 6080-6092, Oct. 2014.

(C8) M. Sawaby, **B. Mamandipoor**, U. Madhow, A. Arbabian, "Analog Processing to Enable Scalable High-Throughput mm-Wave Wireless Fiber Systems" **invited paper**, *50th Asilomar Conference on Signals, Systems and Computers*, Pacific Grove, CA, Nov. 2016.

(C7) **B. Mamandipoor**, M, Fallahpour, G. Malysa, K. Noujeim, A. Arbabian, U. Madhow, "Spatial-Domain Technique to Overcome Grating Lobes in Sparse Monostatic mm-Wave Imaging Systems," *MTT-S International Microwave Symposium (IMS)*, San Francisco, CA, 2016.

(C6) **B. Mamandipoor**, A. Arbabian and U. Madhow, "New Models and Super-Resolution Techniques for Short-Range Radar: Theory and Experiments," **invited paper**, *2016 Information Theory and Applications Workshop (ITA)*, La Jolla, CA, 2016, pp. 1-7.

(C5) **B. Mamandipoor**, M. Sawaby, A. Arbabian, U. Madhow, "Hardware-Constrained Signal Processing for mm-wave LoS MIMO," **invited paper**, *49th Asilomar Conference on Signals, Systems and Computers*, Pacific Grove, CA, Nov. 2015.

(C4) **B. Mamandipoor**, D. Ramasamy, U. Madhow, "Frequency Estimation for a Mixture of Sinusoids: A Near-Optimal Sequential Approach," 3^{rd} *IEEE Global Conference on Signal and Information Processing (GlobalSIP)*, Orlando, FL, Dec. 2015.

(C3) **B. Mamandipoor**, G. Malysa, A. Arbabian, U. Madhow, K. Noujeim, "60 GHz Synthetic Aperture Radar for Short-Range Imaging: Theory and Experiments," *48th Asilomar Conference on Signals, Systems and Computers,* Pacific Grove, CA, Nov. 2014.

(C2) K. Moshksar, **B. Mamandipoor**, Amir K. Khandani, "On Orthogonal Signalling in Gaussian Multiple Access Channel With Peak Constraints," *proceedings of the 2013 IEEE International Symposium on Information Theory (ISIT)*, July 2013, Istanbul, Turkey.

(C1) **B. Mamandipoor**, K. Moshksar, Amir K. Khandani, "On The Sum-Capacity of Gaussian MAC With Peak Constraint," *proceedings of the 2012 IEEE International Symposium on Information Theory (ISIT)*, pp 26-30, July 2012, Boston, MA, USA.

Abstract

Short-Range Millimeter-Wave Sensing and Imaging: Theory, Experiments and Super-Resolution Algorithms

by

Babak Mamandipoor

Recent advancements in silicon technology offer the possibility of realizing low-cost and highly integrated radar sensor and imaging systems in mm-wave (between 30 and 300 GHz) and beyond. Such active short-range mm-wave systems have a wide range of applications including medical imaging, security scanning, autonomous vehicle navigation, and human gesture recognition. Moving to higher frequencies provides us with the spectral and spatial degrees of freedom that we need for high resolution imaging and sensing application. Increased bandwidth availability enhances range resolution by increasing the degrees of freedom in the time-frequency domain. Cross-range resolution is enhanced by the increase in the number of spatial degrees of freedom for a constrained form factor. The focus of this thesis is to explore system design and algorithmic development to utilize the available degrees of freedom in mm-wave frequencies in order to realize imaging and sensing capabilities under cost, complexity and form factor constraints.

We first consider the fundamental problem of estimating frequencies and gains in a noisy mixture of sinusoids. This problem is ubiquitous in radar sensing applications, including target range and velocity estimation using standard radar waveforms (e.g., chirp or stepped frequency continuous wave), and direction of arrival estimation using an array of antenna elements. We have developed a fast and robust iterative algorithm for super-resolving the frequencies and gains, and have demonstrated near-optimal performance in terms of frequency estimation accuracy by benchmarking against the Cramér Rao

Bound in various scenarios.

Next, we explore cross-range radar imaging using an array of antenna elements under severe cost, complexity and form factor constraints. We show that we must account for such constraints in a manner that is quite different from that of conventional radar, and introduce new models and algorithms validated by experimental results. In order to relax the synchronization requirements across multiple transceiver elements we have considered the monostatic architecture in which only the co-located elements are synchronized. We investigate the impact of sparse spatial sampling by reducing the number of array antenna elements, and show that "sparse monostatic" architecture leads to grating lobe artifact, which introduces ambiguity in the detection/estimation of point targets in the scene. At short ranges, however, targets are "low-pass" and contain extended features (consisting of a continuum of points), and are not well-modeled by a small number of point scatterers. We introduce the concept of "spatial aggregation," which provides the flexibility of constructing a dictionary in which each atom corresponds to a collection of point scatterers, and demonstrate its effectiveness in suppressing the grating lobes and preserving the information in the scene.

Finally, we take a more fundamental and systematic approach based on singular decomposition of the imaging system, to understand the information capacity and the limits of performance for various geometries. In general, a scene can be described by an infinite number of independent parameters. However, the number of independent parameters that can be measured through an imaging system (also known as the degrees of freedom of the system) is typically finite, and is constrained by the geometry and wavelength. We introduce a measure to predict the number of spatial degrees of freedom of 1D imaging systems for both monostatic and multistatic array architectures. Our analysis reveals that there is no fundamental benefit in multistatic architecture compared to monostatic in terms of achievable degrees of freedom. The real benefit of multistatic

architecture from a practical point of view, is in being able to design sparse transmit and receive antenna arrays that are capable of achieving the available degrees of freedom. Moreover, our analytical framework opens up new avenues to investigate image formation techniques that aim to reconstruct the reflectivity function of the scene by solving an inverse scattering problem, and provides crucial insights on the achievable resolution.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

Advances in silicon processes and RFIC design imply that there is immense commercial potential for millimeter (mm) wave systems, as long as stringent cost, power and form factor constraints can be satisfied. Mm-wave frequencies provide a significant increase in the number of spatial and spectral degrees of freedom (DoF) compared to legacy narrowband and low frequency systems. In particular, the number of spatial degrees of freedom scale quadratically with the carrier frequency (for a 2D array with fixed form factor). The DoF increase translates to improvement in the achievable range and cross-range resolution for sensing and imaging applications, and for that, the mm-wave band is an attractive candidate: 14 GHz of spectrum (57 to 71 GHz) available at a high carrier frequency on an unlicensed basis [1].

The focus of this thesis is on the system design and algorithmic development of mm-wave radar sensing and imaging systems for applications such as gesture recognition (e.g., implemented on handhelds, laptops or televisions) and vehicular/drone situational awareness [2, 3]. Recently, Google's Project Soli [4] demonstrated a radar sensor chip at 60 GHz frequency, capable of recognizing mm-scale movements of a human hand/fingers, and classifying them in a set of predefined gestures. Authors in [5], have explored the

feasibility of tracking arbitrary movements of a passive device (e.g., a pen) using 60 GHz radar sensors. The goal in [4, 5] is to recognize *moving gestures* by analyzing the *temporal changes* in the back-scattered signal from the scene. However, this is quite different from our focus in this dissertation, which is to extract the information from a *static scene* by reconstructing the spatial configuration of the objects and creating an image.

Radar imagers can be used as an alternative or complementary sensor to optical imagers [6]. Sensing capability of mm-wave arrays can also be used to boost the performance of wireless communication networks. For example, the authors in [7] use the estimates of departure and arrival angles of dominant paths between transmit and receive arrays, to infer the geometrical structure of strong (specular) reflections in the environment. This information can be used for beamforming/nullforming in communication mode. Some of the main advantages of radar systems are direct accessibility to range and Doppler information, the capability of 3D image reconstruction and multi-target tracking applications. Radar imaging systems can also provide the possibility of seeing behind obstacles, i.e., when the desired information is not visually accessible [8] (e.g., seeing behind walls). For example, [9, 10] use WiFi received signal strength (RSS) measurements from moving robotic platforms to image a 3D area behind walls. There are many potential applications for such systems including surveillance, search and rescue operations, and location-aware services [9]. Robustness against weather and lighting conditions and the possibility of electronic beam-steering are other attractive properties of radar imagers.

In this dissertation, we explore fundamental advances in super-resolution and sparse reconstruction algorithms for sensing and imaging applications, and develop a theoretical framework for short-range mm-wave radar, validated by experiments. In the first part of the dissertation (Chapter 2), we focus on frequency estimation from a mixture of sinusoids in additive white Gaussian noise (AWGN), which is a fundamental problem that arises in a variety of communication and radar applications, including estimation of spatial chan-

nels (e.g., for phased arrays), temporal multipath channels (e.g. for equalization), and spatiotemporal channels (e.g., range and direction of arrival estimation for a target). In all of these applications, the spectrum of the measured signal consists of multiple discrete frequencies over a continuous interval. The problem of estimating the frequencies of such a signal is also known as line spectral estimation. We propose an algorithm that generalizes the well-known Orthogonal Matching Pursuit (OMP) to a continuously parametrized over-complete basis using Newton refinements, and is therefore named Newtonized-OMP (NOMP). These refinement phases are crucial for avoiding basis mismatch and obtaining accuracies far better than what would be possible by optimizing over a discrete grid.

In the second part of the dissertation (Chapters 3 and 4), we investigate the design and analysis of imaging systems using an array of radar transceivers that are constrained to be located on a surface with relatively small form factor (e.g., a smart phone or a tablet computer), and the range of interest is of the order of a meter or less. We consider two canonical active imaging array architectures known as *monostatic* and *multistatic* [11], shown in Figure 1.1. A monostatic array consists of standalone transceiver (TRx) elements, i.e., when one of the elements is transmitting, only the co-located receiver collects the back-scattered signal. On the other hand, in a multistatic architecture for any transmitting (Tx) element, all of the receivers (Rx) across the array collect the back-scattered signal. In order to implement a multistatic array, we need to *synchronize* all of the TRx elements across the aperture. Achieving synchronization at mm-wave frequencies and across large baselines is a challenging task and increases the cost and complexity of the system significantly. On the other hand, it is known that multistatic architecture is capable of measuring a greater portion of the k-space spectrum for any point scatterer in the scene. Therefore, it is of significant interest to understand the potential improvements in the information capacity of multistatic array compared to that of monostatic.

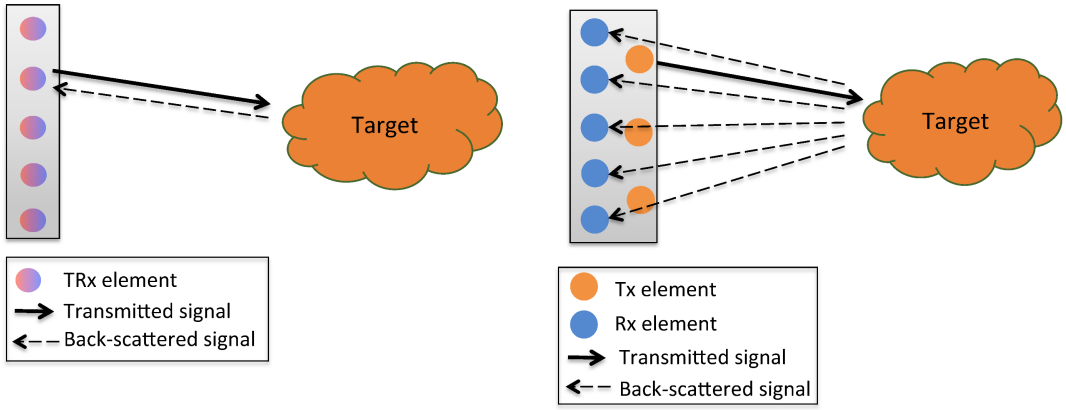


Figure 1.1: Canonical imaging array architectures, (left) monostatic and, (right) multistatic.

The rest of this chapter represents a high level introduction to the problems that have been considered in the dissertation, along with a summary of our main findings. Detailed discussions of the related work to each problem appears in the introduction of each chapter.

1.1 Frequency Estimation in a Mixture of Sinusoids

Chapter 2 of this thesis presents an algorithm to estimate frequencies from N equi-spaced noisy samples in time, denoted by $\mathbf{y} \in \mathbb{C}^N$. Defining the unit norm sinusoid of frequency $\omega \in [0, 2\pi)$, by

$$\mathbf{x}(\omega) \triangleq \frac{1}{\sqrt{N}} \left[1 \; e^{j\omega} \; \cdots \; e^{j(N-1)\omega}\right]^T, \tag{1.1}$$

the observed signal is a mixture of K sinusoids:

$$\mathbf{y} = \sum_{l=1}^{K} g_l \mathbf{x}(\omega_l) + \mathbf{z}, \;\; \mathbf{z} \sim \mathcal{CN}\left(\mathbf{0}, \sigma^2 \mathbb{I}_N\right), \tag{1.2}$$

where $g_l \in \mathbb{C}$ are the unknown complex gains. The signal to noise ratio for l^{th} sinusoid is given by $\text{SNR}_l = |g_l|^2/\sigma^2$. The goal of the algorithm is to provide reliable estimates of $\{(g_l, \omega_l) : l = 1, 2, \ldots, K\}$ and K, the number of sinusoids in the mixture.

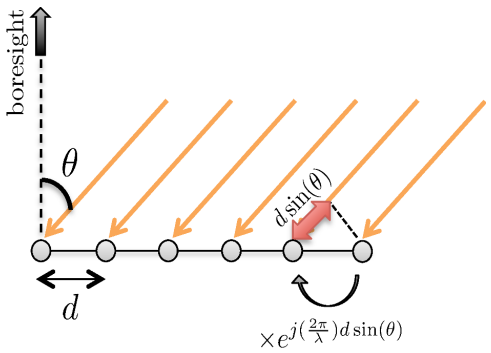


Figure 1.2: Estimating the direction of arrival of an electromagnetic beam impinging on a uniform array of sensors at angle θ. The phase difference between two nearby sensors is given by $\Delta\phi = (\frac{2\pi}{\lambda})d\sin(\theta)$, hence creating a linear phase pattern across the array.

The preceding model and its variants have many applications. For a linear array with N elements with inter-element spacing d, the response corresponding to angle of arrival or departure θ relative to boresight (shown in Figure 1.2) is given by $\mathbf{x}(\omega)$, where $\omega = 2\pi(d/\lambda)\sin(\theta)$ is the spatial frequency corresponding to θ, and λ denotes the carrier wavelength. For estimation of a multipath channel $h(t) = \sum_{l=1}^{K} g_l\delta(t - \tau_l)$, the channel transfer function is given by $H(f) = \sum_{l=1}^{K} g_l e^{-j2\pi f\tau_l}$. Sampling uniformly in the frequency domain with spacing Δf, yields a mixture of sinusoids with $\omega_l = -2\pi\Delta f\tau_l$, reducing the problem of estimating delays to that of frequency estimation. This directly models the operation of stepped frequency continuous wave (SFCW) for imaging a collection of point scatterers. When the channel is "seen through" a pulse $p(t)$, as is often the case for channel estimation in communication applications, then the noiseless frequency domain signal is given by $Y(f) = \sum_{l=1}^{K} g_l P(f)e^{-j2\pi f\tau_l}$. A simple extension of our algorithm to handle weighted sinusoids applies to this setting. We do not provide detailed discussion here, but the code that we have made available [12] does provide the required flexibility.

We are interested in the setting where K is a small integer, so that the underlying signal has a sparse representation in the atomic set of unit norm sinusoids. The goal is to find the best sparse approximation of $\mathbf{y}$ using atoms in the dictionary of sinusoids. A promising algorithm in sparse approximation theory is OMP [13, 14], a greedy algorithm that iteratively identifies the atom that yields the greatest improvement in approximation quality. If the structure of the atomic set is "simple," so that identifying the "best" atom in each iteration needs a small amount of computation, then OMP becomes computationally attractive. Unfortunately, for our atomic set, searching over a continuum of atoms is not possible. An approximation that is often employed to overcome this problem is to discretize the set of frequencies, hoping that the signal still admits a sparse representation in this finite set. However, as discussed in [15], no matter how finely we grid the parameter space, the true underlying atoms of $\mathbf{y}$ need not lie on the grid. This "off-grid effect," or basis mismatch, degrades the performance of reconstruction algorithms significantly. The authors in [16] propose inserting a gradient-based local search in a Matching Pursuit [17] framework, as a means of alleviating the off-grid effect. We go one step further by incorporating a Newton-based Cyclic Refinement step into the OMP framework: this not only sidesteps the off-grid effect, but also enhances performance by refining previously estimated atoms in each iteration.

1.1.1 Contributions

Our key contributions are as follows:

1. We propose Newtonized OMP (NOMP), which detects the best atom over a discrete grid, but avoids basis mismatch by adding a Newton refinement step, thus emulating pursuit over the continuum. In addition, we go beyond OMP by locally

refining all estimated atoms in each iteration, thus re-evaluating estimates of previously detected sinusoids to incorporate the effect of the newly detected sinusoid. This second refinement step can be interpreted as feedback presented to the atoms we have already detected. We show that this feedback mechanism plays a crucial role in handling *interference* among the underlying atoms of $\mathbf{y}$, yielding estimation accuracies far better than what would be possible by standard greedy pursuit algorithms (e.g., OMP and Matching Pursuit). We do not require explicit estimates of model order, and provide a stopping criterion based on CFAR (constant false alarm rate) based on an estimate of the noise variance.

2. We prove convergence of NOMP by providing upper bounds on the number of iterations. Moreover, we derive a bound on the convergence rate, showing that by choosing a fine enough grid for detection, the bound approaches the rate of convergence for OMP on the continuum.

3. We show that the algorithm is near-optimal by numerical comparisons against the Cramér Rao Bound (CRB) [18] in a variety of settings. When the frequencies of the sinusoids in the mixture are well-separated, NOMP is able to achieve the CRB for any sinusoid that has an SNR greater than a certain threshold. Moreover, NOMP is able to resolve closely-spaced frequencies with near-optimal accuracy, as long as there is enough disparity in the SNRs of different sinusoids.

4. We analyze the computational complexity of NOMP. Our numerical experiments show that its run time is significantly smaller than that of recently proposed state-of-the-art Atomic norm Soft Thresholding (AST) algorithm [19], and is even smaller

than that of the classical MUSIC algorithm [20]. We evaluate the run time of the algorithm in a variety of settings. A freely downloadable software package implementing the proposed algorithm can be found in [12].

1.2 Short-range Imaging Under Cost and Complexity Constraints

While the long-range radar systems used in avionics and defense systems today are bulky and expensive, recent advances in silicon offer the possibility of low-cost, highly integrated radar systems at mm-wave and beyond [11, 21]. The latter are particularly well matched to emerging short-range active radar imaging applications such as autonomous vehicular navigation [22], and human gesture recognition on a handheld device [23, 4]. In order to minimize cost and complexity, we wish to use as few elements as possible, while enforcing minimal or no cooperation between the array elements during data acquisition. Specifically, we consider a "sparse monostatic" array architecture in which spatially separated transceiver elements are not synchronized in frequency and phase (and only loosely synchronized in time). The constraints on the spatial extent of the array and the scene, together with the sparsity of the array, imply that the standard framework for long-range radar does not apply, and motivates the effort in Chapter 3 of this thesis to develop new models and algorithms, and to identify fundamental limits imposed by geometric constraints.

1.2.1 Contributions

Our key contributions are as follows:

1. We first show that constraints on array form factor and scene extent limit the

number of spatial DoF. Employing more array elements than the number of degrees of freedom can improve SNR, but does not improve normalized measures of target discrimination.

2. We explore the effect of *sparse arrays* with number of elements significantly smaller than the degrees of freedom. For conventional point scatterer basis functions, it is well known that sparse arrays create grating lobes (i.e., targets in spatially separated locations can have highly correlated responses), thereby creating fundamental ambiguities. At short ranges, however, targets contain extended features (consisting of a continuum of points), and are not well-modeled by a small number of point scatterers. We introduce the concept of "spatial aggregation," which provides the flexibility of constructing a dictionary in which each atom corresponds to a collection of point scatterers. Specifically, we show that patch-based basis functions alleviate the problem of grating lobes with sparse arrays.

3. We show how to choose the patch sizes based on estimation-theoretic bounds: elimination of grating lobes is related to the SNR threshold at which the Ziv-Zakai bound (ZZB) converges to the CRB.

4. We show that, while spatial aggregation helps with standard "matched filter" style imaging, it also provides an effective basis for *sparse representation* of simple scenes. Namely, if the scene reflectivity is lowpass, then patch-based basis functions are able to provide a parsimonious representation of the information in the scene. This sparse representation can be further exploited for image reconstruction in order to super-resolve the scene beyond the capability of conventional methods. We apply NOMP algorithm for sparse image reconstruction, benchmarking against standard ℓ_1-regularized convex optimization. We illustrate the efficacy of our ideas and algorithms using a testbed in which a quasi-monostatic transceiver at 60 GHz

(wavelength of $\lambda = 0.5$ cm) is used to emulate (for static scenes) a two-dimensional (2D) array using a movable platform.

1.3 Geometry-Constrained Degrees of Freedom of Imaging Systems

The evaluation of the amount of information in an unknown scene (object) that can be inferred through measurements of radiated (or back-scattered) electromagnetic fields, is a fundamental problem that has relevance across different fields including optics [24, 25], wireless communications [26, 27], and electromagnetic imaging [28]. One of the crucial measures of the information capacity of such systems is the number of DoF. In general, a scene can be described by an infinite number of independent parameters. However, the number of independent parameters that can be measured through an imaging system is typically finite [29], and is given by the number of DoF of the system. The available DoF is only constrained by the geometry and wavelength (i.e., we assume no limitation on the number of array elements over the aperture). We formulate the problem under "Born approximation" [30], that is under the assumption of *weak scattering* model where the total electromagnetic field at the scene is approximated by the incident field. Under this assumption, the measurement model is linearized, hence, we can resort to singular value decomposition (SVD) to analyze the model. Additionally, we present a theory for DoF evaluation of narrowband (single frequency) 1D monostatic and multistatic imaging systems, and provide guidelines for system design and performance evaluation of image reconstruction techniques.

1.3.1 Contributions

Our key contributions are as follows:

1. We introduce *space-bandwidth product* (SBP), defined by the product of the scene area and the measured Fourier extent of the scene (after removing redundancies), as a means of identifying the DoF of the system. SBP can be thought of as a generalization of the so-called Shannon number [31], for spectral measurements of an unknown scene through a space-variant bandlimitted system. We evaluate the accuracy of our proposed DoF measure by comparing it to numerical SVD computations for various geometries.

2. We investigate the SBP analysis in the Fresnel zone, and deduce the approximate answers of previous models that were based on Fresnel approximation between parallel planar surfaces. We also provide a clear analysis of the design guidelines for multistatic arrays that are based on the "effective aperture" concept.

3. We investigate image formation techniques that aim to reconstruct the reflectivity function of the scene by solving an inverse scattering problem. We show that the SVD analysis provides an easy understanding of the measurement process, as well as the achievable resolution of various reconstruction schemes. In particular, we show that back-propagation reconstruction for multitstatic architecture is highly suboptimal and leads to a significant resolution loss in the image.

Chapter 2

Newtonized Orthogonal Matching Pursuit (NOMP) Algorithm

2.1 Introduction

We propose an algorithm to estimate frequencies from N equi-spaced noisy samples in time, denoted by $\mathbf{y} \in \mathbb{C}^N$. Defining the unit norm sinusoid of frequency $\omega \in [0, 2\pi)$, by

$$\mathbf{x}(\omega) \triangleq \frac{1}{\sqrt{N}} \left[1 \; e^{j\omega} \; \cdots \; e^{j(N-1)\omega}\right]^T, \tag{2.1}$$

the observed signal is a mixture of K sinusoids:

$$\mathbf{y} = \sum_{l=1}^{K} g_l \mathbf{x}(\omega_l) + \mathbf{z}, \; \mathbf{z} \sim \mathcal{CN}\left(\mathbf{0}, \sigma^2 \mathbb{I}_N\right), \tag{2.2}$$

where $g_l \in \mathbb{C}$ are the unknown complex gains. The signal to noise ratio for l^{th} sinusoid is given by $\text{SNR}_l = |g_l|^2/\sigma^2$. Our goal is to provide reliable estimates of $\{(g_l, \omega_l) : l = 1, 2, \ldots, K\}$ and K, the number of sinusoids in the mixture.

Parts of this chapter are reprinted from our journal paper [32], with permission

Our proposed algorithm is a natural generalization of OMP to the continuum using Newton refinements, and hence is termed Newtonized OMP (NOMP). Each iteration consists of two phases: detection of a new sinusoid, and sequential Newton refinements of the parameters of already detected sinusoids. The refinements play a critical role in two ways: (1) sidestepping the potential *basis mismatch* from discretizing a continuous parameter space, (2) providing *feedback* for locally refining parameters estimated in previous iterations. We characterize convergence, and provide a Constant False Alarm Rate (CFAR) based termination criterion. By benchmarking against the Cramér Rao Bound, we show that NOMP achieves near-optimal performance under a variety of conditions. We compare the performance of NOMP with classical algorithms such as MUSIC and more recent Atomic norm Soft Thresholding (AST) and Lasso algorithms, both in terms of frequency estimation accuracy and run time.

2.1.1 Related work

Line spectral estimation is a fundamental problem in statistical signal processing. Classical (and popular) subspace methods such as MUSIC and ESPRIT [20, 33] exploit the low-rank structure of the autocorrelation matrix. One of the major advantages of these methods is the capability of resolving multiple closely-spaced frequencies at high SNR. Both MUSIC and ESPRIT have been shown to be asymptotically optimal in the limit of infinite SNR [34], but their performance degrades at medium and low SNRs. Another family of DFT-based classical methods [35, 36], typically have lower computational complexity and estimation accuracy similar to that of subspace methods [36, 37].

More recent techniques using convex optimization cast the frequency estimation problem as that of finding a sparse approximation of the received signal using an infinite-dimensional dictionary of sinusoids. It is shown in [38] that, in the absence of noise,

total-variation norm is able to locate frequencies with infinite precision, as long as the minimum frequency separation exceeds $4\times\Delta_{\text{dft}}$ where $\Delta_{\text{dft}} = 2\pi/N$ is the Discrete Fourier Transform (DFT) grid spacing. The sufficient condition on required minimum separation has been recently improved to $2.52 \times \Delta_{\text{dft}}$ in [39]. An extension to noisy scenarios is provided in [40]. Another approach is AST [19], [41], which provides theoretical guarantees of noise robustness in terms of mean squared error (MSE). Both total-variation norm and atomic norm are generalizations of the ℓ_1 norm to infinite-dimensional settings. Solving these optimization problems involves solving the Lagrange dual which takes the form of a semi-infinite program (SIP) with finite-dimensional decision variables and infinitely many constraints. For the sparse frequency estimation problem, [40] and [19] reformulate the dual as a semidefinite program (SDP), which enables numerical optimization. Similar reformulation for other problems seems to be difficult [41]. A pragmatic approach is to use Lasso optimization on a highly oversampled grid as an approximation for AST [19, 41]. Both AST and Lasso are benchmarks that we compare our proposed algorithm against in our numerical experiments.

NOMP can be viewed as coordinate optimization over a continuum. Coordinate-wise descent has been widely used for sparse approximation; for example, such methods are competitive for solving Lasso type problems [42, 43]. Preliminary results in [44] show that coordinate descent to a relaxation of the AST problem can be a means to speed up implementation.

There is a large body of work on feedback for improving the performance of iterative greedy algorithms for sparse approximation [45], [46], [47], [48], [49], [50]: information from recent iterations is used to remove errors introduced by previous steps. For example, [46] introduces a forward-backward greedy algorithm that allows *aggressive* backward steps (discarding small-magnitude atoms) after a greedy forward step. In contrast, NOMP, by virtue of its continuous parameterization, can employ a *mild* form of feedback

by locally refining the set of estimated atoms, thus allowing an atom to be replaced by a nearby, highly correlated atom, in the continuum. The idea of embedding a local refinement step in OMP has been proposed in [51] in a discrete setting, but this algorithm does not address basis mismatch, and assumes that the model order is known *a priori.*

Many theoretical results on OMP [14, 52, 53] are based on assumptions such as incoherence or Restricted Isometry Property (RIP) for the underlying dictionary. Such approaches do not work for analyzing NOMP, since nearby atoms are highly correlated for continuous frequency estimation. Instead, our convergence analysis borrows tools from analysis of AST-based line spectral estimation [19], along with observations following from our CFAR-based design.

For a single sinusoid, frequency estimation using coarse detection followed by Newton refinement was proposed three decades ago by Abatzoglou [54]. This was recently adapted for estimation of a single delay in [55], and shown to approach estimation-theoretic bounds. In prior work, sequential algorithms similar to NOMP have been used for millimeter wave spatial channel estimation with compressive measurements [56, 57]. The NOMP algorithm presented here is an improvement on these, with a principled CFAR-based stopping criterion. We present NOMP within an application-independent abstraction, recognizing the fundamental nature and widespread utility of the frequency estimation problem. We have reported on a version of NOMP in a conference paper [58] (although the algorithm in [58] omits a least squares step included here for establishing convergence rate results, but with little impact on practical performance). The algorithm presented in this dissertation has appeared in our journal paper [32].

A closely related algorithm to NOMP is proposed in [59], which employs a Bayesian framework for frequency estimation, using Newton refinements for updating the frequencies. The details are different from our non-Bayesian, CFAR framework, and convergence analysis is not provided, but the benefits of Newtonization are also evident in the nu-

merical results in [59].

Map of this Chapter: We present our algorithm in Section 2.2. In Section 2.3, we present the CFAR-based stopping criterion, together with a simplified analytical characterization of false alarm and miss probabilities. In Section 2.4, we show why oversampling is essential for discretization followed by refinement to emulate pursuit over the continuum. We discuss convergence in Section 2.5. In Section 2.6, we report on numerical experiments, and compare NOMP with other methods in terms of estimation accuracy and computational complexity. Section 2.7 discusses extension of the algorithm to more general settings, with illustrative numerical results for compressive measurements. We set $N = 256$ throughout in our numerical results (other values of N yield entirely similar trends).

Notation: Complex conjugate transpose of v is denoted by v^H. $\Re\{a\}$ is the real part of complex number a. The Moore-Penrose pseudo-inverse of matrix A is denoted by $A^\dagger$. The distance between any two frequencies $\{\omega_l, \omega_k\}$ is defined by $\text{dist}(\omega_k, \omega_l) \triangleq \min_{a\in\mathbb{Z}} |\omega_k - \omega_l + 2\pi a|$, i.e. the wrap-around distance when we restrict frequencies to lie in $[0, 2\pi)$. The DFT matrix with unit norm columns and the corresponding grid spacing are denoted by $\mathcal{F}$ and Δ_{dft}, respectively. The inner product between $v, u \in \mathbb{C}^N$ is defined as $\langle v, u\rangle = u^H v$.

2.2 NOMP Algorithm

We first discuss estimation of a single sinusoid, and then build on it to generalize to a mixture of sinusoids.

2.2.1 Single frequency

We have $\mathbf{y} = g\mathbf{x}(\omega) + \mathbf{z}$. The Maximum Likelihood (ML) estimate of the gain and frequency are obtained by minimizing the residual power $\|\mathbf{y} - g\mathbf{x}(\omega)\|^2$, or equivalently, by maximizing the function

$$S(g, \omega) = 2\Re\{\mathbf{y}^H g\mathbf{x}(\omega)\} - |g|^2 \|\mathbf{x}(\omega)\|^2 . \tag{2.3}$$

Directly optimizing $S(g, \omega)$ over all gains and frequencies is difficult. Therefore, we adopt a two stage procedure: (1) Detection stage, where we find a coarse estimate of ω by restricting it to a discrete set, (2) Refinement stage, in which we iteratively refine gain and frequency estimates.

For any given ω, the gain that maximizes $S(g, \omega)$ is given by $\hat{g} = (\mathbf{x}(\omega)^H \mathbf{y}) / \|\mathbf{x}(\omega)\|^2$. Substituting $\hat{g}$ in $S(g, \omega)$ yields that the generalized likelihood ratio test (GLRT) estimate of ω (treating g as a nuisance parameter) is the solution to the following optimization problem:

$$\hat{\omega} = \arg\max_{\omega} \; G_{\mathbf{y}}(\omega), \tag{2.4}$$

where

$$G_{\mathbf{y}}(\omega) = |\mathbf{x}(\omega)^H \mathbf{y}|^2 / \|\mathbf{x}(\omega)\|^2 \tag{2.5}$$

is the GLRT cost function. We use this observation to find a coarse estimate of (g, ω) in the Detection stage.

Detection: We obtain a coarse estimate of ω by restricting it to a finite discrete set denoted by $\Omega \triangleq \{k(2\pi/\gamma N) : k = 0, 1, \ldots, (\gamma N - 1)\}$, where γ is the over-sampling factor relative to the DFT grid. For our simulation results, we set $\gamma = 4$. The outputs of this stage are $\omega_{\mathrm{c}} \in \Omega$ that maximizes the GLRT cost function (2.4), and the corresponding gain $(\mathbf{x}(\omega_{\mathrm{c}})^H \mathbf{y}) / \|\mathbf{x}(\omega_{\mathrm{c}})\|^2$.

Refinement: Since ω can take any value in interval $[0, 2\pi)$, we add a Newton-based refinement stage for estimation on the continuum. Let $(\hat{g}, \hat{\omega})$ denote the current estimate. The Newton step for frequency refinement is given by

$$\hat{\omega}' = \hat{\omega} - \dot{S}(\hat{g}, \hat{\omega}) / \ddot{S}(\hat{g}, \hat{\omega}) \tag{2.6}$$

where

$$\begin{aligned} \dot{S}(g, \omega) &= \Re\{(\mathbf{y} - g\mathbf{x}(\omega))^H g(d\mathbf{x}(\omega)/d\omega)\} & (2.7) \\ \ddot{S}(g, \omega) &= \Re\{(\mathbf{y} - g\mathbf{x}(\omega))^H g(d^2\mathbf{x}(\omega)/d\omega^2)\} & (2.8) \\ &\quad -|g|^2 \left\| d\mathbf{x}(\omega)/d\omega \right\|^2 . \end{aligned}$$

As we want to *maximize* $S(g, \omega)$, we only apply the update rule (2.6) when the function is locally concave (i.e. $\ddot{S}(\hat{g}, \hat{\omega}) < 0$). The gain parameter is then updated to maximize $S(g, \hat{\omega}')$: $\hat{g}' = (\mathbf{x}(\hat{\omega}')^H \mathbf{y}) / \left\| \mathbf{x}(\hat{\omega}') \right\|^2$.

Refinement Acceptance Condition (`RAC`): We *accept* a refinement only if it leads to a strict improvement in $G_{\mathbf{y}}(\omega)$; that is, if $G_{\mathbf{y}}(\hat{\omega}') > G_{\mathbf{y}}(\hat{\omega})$. This ensures that an accepted refinement can only decrease the overall residual energy, and that the residual energy is non-increasing throughout the course of the algorithm, which ensures convergence, as shown in Section 2.5.

2.2.2 Multiple frequencies

Let $\mathcal{P} = \{(g_l, \omega_l)\,,\ l = 1, \ldots, k\}$ denote a set of estimates of the parameters of the sinusoids in the mixture. Let

$$\mathbf{y}_{\mathrm{r}}(\mathcal{P}) = \mathbf{y} - \sum_{l=1}^{l=k} g_l \mathbf{x}(\omega_l) \tag{2.9}$$

denote the *residual* measurement corresponding to this estimate. The following procedure is a direct generalization of the single sinusoid refinement algorithm to multiple frequencies. It proceeds by employing the single sinusoid algorithm to perform *Newtonized coordinate descent* on the overall residual energy $\|\mathbf{y}_{\mathrm{r}}(\mathcal{P})\|^2$. One step of this coordinate descent involves cycling through all sinusoids in $\mathcal{P}$ in a predetermined order. In this process, suppose that we wish to refine the l-th sinusoid: we treat $\mathbf{y}_{\mathrm{r}}(\mathcal{P}\setminus\{(g_l,\omega_l)\})$ as our measurement $\mathbf{y}$ and employ the single frequency update step to refine (g_l,ω_l).

Algorithm 1 Newtonized Orthogonal Matching Pursuit

Procedure EXTRACTSPECTRUM($\mathbf{y}$, τ):
$m \leftarrow 0$, $\mathcal{P}_0 = \{\}$
while $\max_{\omega\in\mathrm{DFT}} G_{\mathbf{y}_{\mathrm{r}}(\mathcal{P}_m)}(\omega) > \tau$ **do**
  $m \leftarrow m+1$
  IDENTIFY
  $\hat{\omega} = \arg\max_{\omega\in\Omega} G_{\mathbf{y}_{\mathrm{r}}(\mathcal{P}_{m-1})}(\omega)$
  and its corresponding gain
  $\hat{g} \leftarrow \left(\mathbf{x}(\hat{\omega})^H \mathbf{y}_{\mathrm{r}}\left(\mathcal{P}_{m-1}\right)\right)/\|\mathbf{x}(\hat{\omega})\|^2$
  $\mathcal{P}'_m \leftarrow \mathcal{P}_{m-1}\cup\{(\hat{g},\hat{\omega})\}$
  SINGLE REFINEMENT: Refine $(\hat{g},\hat{\omega})$ using single frequency Newton update algorithm (R_s Newton steps) to obtain improved estimates $(\hat{g}',\hat{\omega}')$.
  $\mathcal{P}''_m \leftarrow \mathcal{P}_{m-1}\cup\{(\hat{g}',\hat{\omega}')\}$
  CYCLIC REFINEMENT: Refine parameters in $\mathcal{P}''_m$ one at a time: For each $(g,\omega)\in\mathcal{P}''_m$ we treat $\mathbf{y}_{\mathrm{r}}(\mathcal{P}''_m\setminus\{(g,\omega)\})$ as the measurement $\mathbf{y}$, and apply single frequency Newton update algorithm. We perform R_c rounds of cyclic refinements. Let $\mathcal{P}'''_m$ denote the new set of parameters.
  UPDATE all gains in $\mathcal{P}'''_m$ by least squares:
  $X \triangleq [\mathbf{x}(\omega_1)\ldots\mathbf{x}(\omega_m)]$, $\{\omega_l\}$ are the frequencies in $\mathcal{P}'''_m$
  $[g_1\ldots g_m]^T = X^{\dagger}\mathbf{y}$
  Let $\mathcal{P}_m$ denote the new set of parameters.
return $\mathcal{P}_m$

The NOMP procedure is summarized in Algorithm 1. We now briefly discuss the role of its main components:

• SINGLE REFINEMENT (Step 7): Ideally, we want to identify $\hat{\omega}$ which maximizes the GLRT cost function over the continuum. The SINGLE REFINEMENT step emulates search over the continuum by locally refining the estimate of $\hat{\omega}$ obtained by picking the maximum over the discrete set Ω.

• CYCLIC REFINEMENT (Step 9) is where NOMP diverges from *forward greedy methods* [60] (in particular OMP) by providing *feedback* for local refinements of previously detected sinusoids. This gives them an opportunity to better explain the received signal in light of new information regarding the presence of another sinusoid. This feedback is presented in the form of an updated residue. As we see in Sections 2.5.3 and 2.6, this step is crucial for fast convergence and high estimation accuracy.

• UPDATE by least squares (Step 10): Here we update gains by projecting the received signal $\mathbf{y}$ onto the subspace spanned by the estimated frequencies. This ensures that the residual energy is the minimum possible for the current set of estimated frequencies. We see in Section 2.5 that performing this projection step just prior to detecting a new sinusoid enables us to lower bound the convergence rate of NOMP by mirroring arguments used to establish bounds on OMP convergence [60].

We have left the number of refinement steps unspecified so far. For the simulations in this chapter, we set $R_s = 1$ for every newly detected sinusoid in the SINGLE REFINEMENT step, and $R_c \in \{1, 3, 5\}$ refinement rounds in the CYCLIC REFINEMENT step, depending on the difficulty of the estimation problem.

2.2.3 Complexity analysis

We analyze the computational complexity of the main steps of NOMP assuming that the algorithm runs for exactly K iterations (i.e., perfect stopping). Checking whether the stopping criterion is satisfied is efficiently implemented using Fast Fourier Transform

(FFT), with complexity $\mathcal{O}(KN\log(N))$. The IDENTIFY Step involves computing the GLRT cost function over the set of frequencies defined by Ω. This can also be computed using FFTs in $\mathcal{O}(\gamma KN\log(\gamma N))$ time. The SINGLE REFINEMENT Step takes only $\mathcal{O}(R_sN)$ operations per sinusoid, hence the total cost for SINGLE REFINEMENT is $\mathcal{O}(R_sKN)$. The CYCLIC REFINEMENT involves refining all frequencies that have been estimated so far, and has overall complexity $\mathcal{O}(R_cR_sK^2N)$. If we directly compute the pseudo-inverse and apply it to the vector of observation (i.e., $X^\dagger\mathbf{y}$) in the UPDATE Step, the complexity is $\mathcal{O}(NK^2+K^3)$ per iteration, and the overall cost is $\mathcal{O}(NK^3+K^4)$. However, we note that iterative methods (such as Richardson's iteration or conjugate gradient [45]) are extremely efficient in computing the least squares solution and can be used for speeding up the UPDATE Step. Empirical observations suggest that the CYCLIC REFINEMENT step dominates the overall computational cost of NOMP.

Minimum frequency separation: When two frequencies, say ω_1 and ω_2, are "very close", intuitively, the mixture $g_1\mathbf{x}(\omega_1)+g_2\mathbf{x}(\omega_2)$ is explained "very well" by a single frequency, say as $(g_1+g_2)\mathbf{x}(\omega_1)$. Thus, a natural metric to characterize regimes for testing algorithms for mixture frequency estimation is the *minimum frequency separation* between any two sinusoids. We denote this by $\Delta\omega_{\min}=\min_{k\neq l}\operatorname{dist}(\omega_k,\omega_l)$ and we would like our algorithms to work well even for small values of $\Delta\omega_{\min}$.

Without a minimum frequency separation condition, the estimation problem can be hopelessly ill-posed. This has been studied in detail in [38] using Slepian's work on prolate spheroidal sequences [61]. It has been shown that if the frequencies are clustered together, it becomes impossible for *any* method to recover the information from the *noisy* observations. It is important to note that in the limit of infinite SNR, however, one can still estimate a sparse clustered set of frequencies regardless of their separation (e.g., using Prony's method of polynomial interpolation [62]).

Estimation Theoretic Bounds: Estimation theoretic quantities such as Cramér Rao Bound (CRB) and Ziv-Zakai Bound (ZZB) [63] provide lower bounds on the variance of estimators. For single frequency estimation, these bounds are given by [64]

$$\mathrm{CRB}(SNR) = \frac{6}{SNR \times (N^2 - 1)}, \tag{2.10}$$

and

$$\mathrm{ZZB}(SNR) = \int_0^{\pi} Q\left(\sqrt{SNR\left(1 - \left|\frac{\sin(Nh/2)}{N\sin(h/2)}\right|\right)}\right) h \; dh. \tag{2.11}$$

In Figure 2.1, we plot estimation error bounds as a function of $SNR = \|\mathbf{x}(\omega)\|^2/\sigma^2 = 1/\sigma^2$. The ZZB has a distinct threshold behavior: large frequency estimation errors are inevitable in a "low SNR" regime below a threshold, while the ZZB converges to the CRB when the SNR is large enough compared to this threshold. In Section 2.6, we compare algorithms in terms of high SNR behavior relative to the CRB, and low SNR behavior relative to the ZZB threshold. Note that even if per sinusoid SNRs are higher than the ZZB threshold, the joint estimation problem for multiple sinusoids may still be ill-posed (e.g., if the frequency separation is too small).

Remark 1 *We have defined "integrated SNR" obtained by dividing the total power of a sinusoid by the noise power per complex dimension, i.e.,* $SNR = \frac{\mathbb{E}\|\mathbf{x}(\omega)\|^2}{\sigma^2}$. *An alternative definition of signal to noise ratio is the "per-sample SNR", given by* $SNR_{sample} = \frac{\mathbb{E}[\mathbf{x}_i(\omega)]^2}{\mathbb{E}[\mathbf{z}_i]^2} = \frac{1}{N}SNR$. *For example,* $SNR = 25$ *dB (which is the nominal* SNR *value in our simulations), corresponds to* $SNR_{sample} \approx 1$ *dB.*

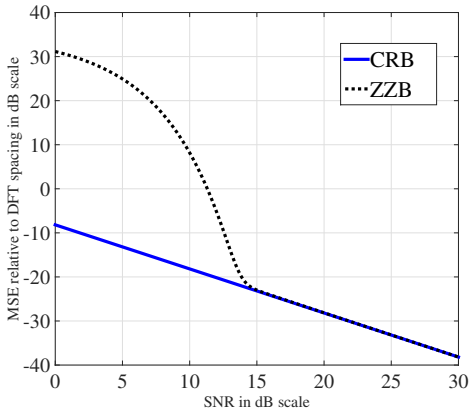


Figure 2.1: CRB and ZZB for estimating the frequency of a single sinusoid.

2.3 CFAR-based Stopping Criterion

Detection problems face a tension between false alarm and missed detection (or miss, for short). In many detection problems, a model for the signal can be elusive. Thus, a common strategy is based on the Constant False Alarm Rate (CFAR) criterion [65, 66], which only requires a model for noise. Here, we use the CFAR criterion to estimate model order (i.e. number of sinusoids in the mixture K): if the residual signal can be explained "well enough" by noise, up to a target false alarm rate, then we stop. We show by simulation that the actual false alarm rate is close to the nominal being designed for.

We also estimate the probability of miss, taking into account the effect of noise but ignoring "interference" from other sinusoids. The resulting receiver operating characteristic (ROC) turns out to be in remarkable agreement with simulations. This shows that, when the sinusoids are separated beyond a minimum separation, and when their SNRs exceed the ZZB thresholds for individual sinusoids, then the probabilities of false alarm and miss are dominated by noise rather than by inter-sinusoid interference.

2.3.1 Stopping criterion

The algorithm terminates when

$$G_{\mathbf{y}_{\mathrm{r}}(\mathcal{P})}(\omega) = \left|\langle \mathbf{y}_{\mathrm{r}}(\mathcal{P}), \mathbf{x}(\omega)\rangle\right|^2 < \tau$$

for all DFT frequencies $\{2\pi k/N \ : \ k = 0, \ldots, N-1\}$. In other words, we stop when $\|\mathcal{F}\mathbf{y}_{\mathrm{r}}(\mathcal{P})\|_\infty^2 < \tau$, where $\mathcal{F}\mathbf{a}$ is the Discrete Fourier Transform of $\mathbf{a}$, and report $\mathcal{P}$ as our estimate of the sinusoids in the mixture.

Suppose that we have already correctly detected all sinusoids in the mixture. In this case, the residual is $\mathbf{y}_{\mathrm{r}}(\mathcal{P}) \approx \mathbf{z}$, where $\mathbf{z} \sim \mathcal{CN}(\mathbf{0}, \sigma^2\mathbb{I}_N)$ (since $\mathcal{F}$ is a projection matrix, the statistics of WGN are unchanged by it). It is easy to show that

$$\Pr\left\{\|\mathcal{F}\mathbf{y}_{\mathrm{r}}(\mathcal{P})\|_\infty^2 > \tau\right\} = 1 - \left(1 - \exp(-\tau/\sigma^2)\right)^N. \tag{2.12}$$

We choose our stopping criterion threshold τ so that $\Pr\left\{\|\mathcal{F}\mathbf{y}_{\mathrm{r}}(\mathcal{P})\|_\infty^2 > \tau\right\} = P_{\mathrm{fa}}$, where P_{fa} is a nominal false alarm rate. Using (2.12), we can explicitly compute this threshold as

$$\tau = -\sigma^2 \log\left(1 - (1 - P_{\mathrm{fa}})^{1/N}\right).$$

A more easily interpreted expression can be obtained via asymptotics for large N [67] (which provide excellent approximations for the moderate values of N used in our numerical results). Let $M_N \triangleq ||\mathcal{F}\mathbf{y}_{\mathrm{r}}(\mathcal{P})||_\infty^2$. If $\mathbf{y}_{\mathrm{r}}(\mathcal{P}) \approx \mathbf{z}$, we have $\mathbb{E}[M_N] = \sigma^2 \sum_{k=1}^N \frac{1}{k} \approx \sigma^2 \log N$, and the asymptotic distribution of $E \triangleq M_N - \sigma^2 \log N$ is given by $\Pr\{E \leq x\} = \exp(-\exp(-x/\sigma^2))$. We set $\tau = \sigma^2 \log(N) + x$, for x so that $\Pr\{E \geq x\}$ is equal to the nominal false alarm rate P_{fa}. This is given by $x = -\sigma^2 \log\log\left(1/(1 - P_{\mathrm{fa}})\right)$ and

the resulting expression for the threshold τ is

$$\tau = \sigma^2 \log(N) - \sigma^2 \log\log\left(1/(1 - P_{\text{fa}})\right). \tag{2.13}$$

The "measured" versus "nominal" false alarm rates for different values of nominal P_{fa}, are shown in Figure 2.2. Each point in the plot is generated by 300 runs of NOMP algorithm for estimating frequencies in a mixture of $K = 16$ sinusoids of fixed nominal SNR. The minimum frequency separation $\Delta\omega_{\min} = 2.5\Delta_{\text{dft}}$. We declare a false alarm whenever NOMP overestimates the model order K. As shown in Figure 2.2, the empirical false alarm rate closely follows the nominal value at various SNRs.

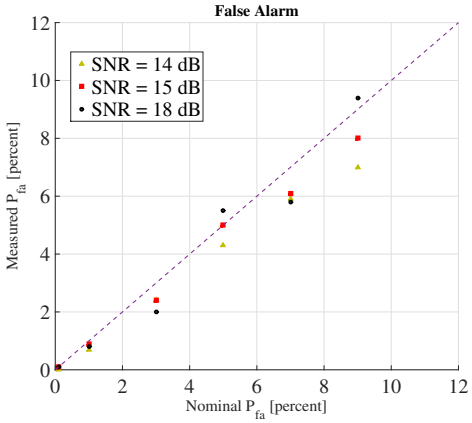


Figure 2.2: Nominal v.s. measured probability of false alarm.

2.3.2 Probability of miss

Define neighborhood N_{ω_i} around each true frequency ω_i by $N_{\omega_i} \triangleq \{\omega : \text{dist}(\omega, \omega_i) < 0.25 \times \Delta_{\text{dft}}\}$. We declare successful detection of ω_i if at least one of the estimated frequencies lies in N_{ω_i}, otherwise we declare a miss for ω_i. A miss results from noise and inter-sinusoid interference, but we only model noise here. For the minimum separation considered here, we show using simulations that the empirical probability of miss is only

a little higher than the analytical estimate that we derive below.

A sinusoid of amplitude A leads to a maximum FFT value of αA, where $\alpha \in [0, 1]$ captures the amplitude reduction due to the grid mismatch. The magnitude of the maximum FFT coefficient, denoted by M_{fft}, is Rician with $\Pr\{M_{\text{fft}} < x\} = 1 - Q_1\left(\frac{\sqrt{2}\alpha A}{\sigma}, \frac{\sqrt{2}x}{\sigma}\right)$, where Q_1 is Marcum Q-function. The sinusoid is not detected by the algorithm if $M_{\text{fft}} < \sqrt{\tau}$, hence

$$P_{\text{miss}} = 1 - Q_1\left(\alpha\sqrt{2SNR}, \sqrt{2\tau/\sigma^2}\right). \tag{2.14}$$

Assuming uniform distribution for a frequency within a DFT grid interval gives us $\mathbb{E}[\alpha] = \mathbb{E}[\sin(N\omega/2)/(N\sin(\omega/2))] = 0.88$, where $\omega \sim \text{Uniform}[-\pi/N, \pi/N]$. Therefore,

$$P_{\text{miss}} \approx 1 - Q_1\left(0.88\sqrt{2SNR}, \sqrt{2\tau/\sigma^2}\right). \tag{2.15}$$

Putting equations (2.13) and (2.15) together, we can characterize the ROC at various SNRs as shown in Figure 2.3. The simulation parameters are the same as Figure 2.2 .We see that for SNR= 18 dB the probability of miss is negligible. When SNR goes below the ZZB threshold e.g. SNR= 14 dB, the probability of miss is bounded away from zero. This behavior is predicted by the ZZB threshold and our ROC analysis for single frequency estimation. However, as we see in Figure 2.3, they serve as excellent approximations for multiple frequency estimation.

Remark 2 *For the simulations in this chapter, we have used the CFAR-based stopping criteria for the NOMP algorithm. However, a variety of other stopping rules, e.g., Bayesian Information Criteria (BIC) [68] and Akaike Information Criteria (AIC) [69], can be easily adapted for use with the NOMP algorithm. In Section 2.6.4, we investigate the performance of the BIC stopping rule (see Appendix I for a quick overview) as well as CFAR-based stopping criteria for NOMP.*

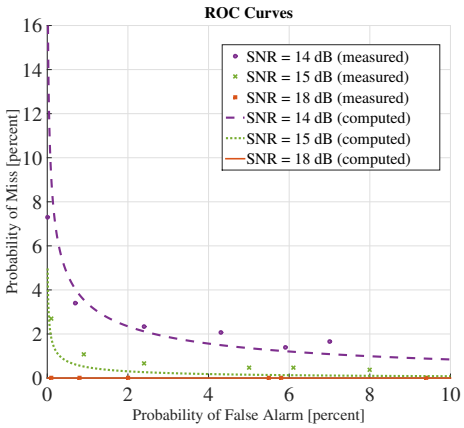


Figure 2.3: Measured v.s. computed ROCs

2.4 The Need to Oversample

In this section, we show that oversampling is indeed required at the detection stage, in order for Newton refinements to converge to the maximum of the GLRT cost function. We ignore noise in these discussions. The GLRT cost function is given by $G_{\mathbf{y}}(\omega) = |\sum_{l=1}^{K} g_l h(\omega - \omega_l)|^2$, where $h(\omega) = \frac{\sin(N\omega/2)}{N\sin(\omega/2)}$ is the Dirichlet kernel. Characterizing the minimum required oversampling factor for this general setting is difficult because of the highly non-convex nature of $G_{\mathbf{y}}(\omega)$. Hence, we focus on a single frequency setting $(K = 1)$.

We wish to arrive at ω_1 by optimizing $G_{\mathbf{y}}(\omega)$ using Newton refinements starting on a coarse grid. We note that $\arg\max_\omega G_{\mathbf{y}}(\omega) = \arg\max_\omega |g_1 h(\omega - \omega_1)|^2 = \arg\max_\omega |h(\omega - \omega_1)|^2$. We would like to characterize the minimum oversampling factor such that if we start off from the best guess of the maximum of $G_{\mathbf{y}_\mathrm{r}}(\omega)$ on the grid, the Newton refinement stage will take us to ω_1. That is, we must characterize how close to ω_1 must the nearest grid point lie, so that Newton refinements will always take us to ω_1 from this grid point. Without loss of generality, we set $\omega_1 = 0$ (since we have shifted our frequency axis such that $\omega_1 = 0$, no grid point may lie on 0).

We start by normalizing frequencies by the DFT spacing. In this scaled frequency axis, the Dirchelet kernel is given by $h(x) = \frac{\sin(\pi x)}{N \sin(\pi x/N)}$, where $x = \omega/\Delta_{\text{dft}}$ is sometimes referred to as the normalized frequency. As shown in [70], the Newton method converges to the solution of $h'(x) = 0$ quadratically, if the initial guess x_0 lies in an interval I around the true solution where the following conditions are met:

- $h''(x) \neq 0 \quad \forall x \in I$.
- $h'''(x)$ is finite $\forall x \in I$.
- $|x_0| < 1/M$ where $M \triangleq \sup_{x \in I} 0.5 \left| \frac{h'''(x)}{h''(x)} \right|$.

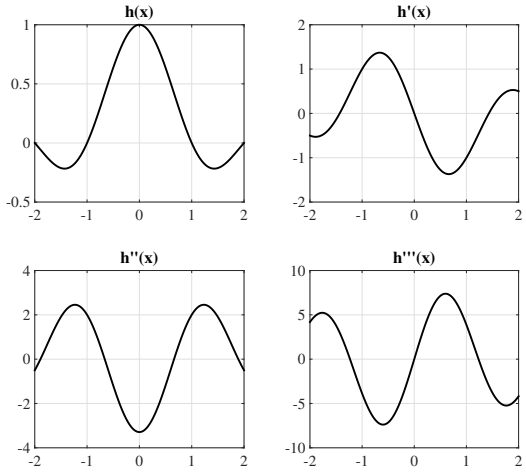


Figure 2.4: Dirichlet kernel h(x) and its derivatives.

Figure 2.4 shows the function $h(x)$ and its derivatives in a window around the origin. The first two conditions are met for any interval $I \subset (-1, 1)$. Therefore, it only remains to satisfy the third condition. By simple algebra one can see that if we set $I = (-0.45, +0.45)$, then we have $|x_0| < 1/M$ for any $x_0 \in I$. Therefore the maximum acceptable grid spacing is about 0.9 which is equivalent to minimum oversampling factor ≈ 1.12. This simple analysis shows that, even for a single sinusoid and no noise, we must

sample beyond the DFT grid to ensure that the two-stage detection/refinement procedure successfully identifies the maximum of the GLRT cost function, thereby imitating pursuit over the continuum. Our simulations show that setting the oversampling factor at $\gamma = 4$ or more works very well, independent of the number of the sinusoids in the observations.

2.5 Convergence

In this section, we first characterize convergence by providing upper bounds on the number of iterations of NOMP. We then provide a bound on the rate of convergence of NOMP as a function of the "atomic norm" of $\mathbf{y}$, and the oversampling factor γ. Our convergence results are pessimistic, in that they do not account for the effect of the refinement steps. We show by simulations the dramatic improvement due to refinements, by comparing NOMP to the following variants of OMP:

Discretized OMP (DOMP): This is standard OMP applied to the oversampled grid of sinusoids Ω. We use our NOMP implementation with number of refinement steps set to zero, increasing the oversampling factor to $\gamma = 20$. Since DOMP can be interpreted as a special case of NOMP, the convergence analysis presented here is valid for DOMP as well.

NOMP without Cyclic Refinements (NOMP–): If we skip the Cyclic Refinement step of NOMP, then we get an algorithm that emulates OMP over the continuum of atoms. Note that NOMP– does not have a feedback mechanism, hence it lies in the class of forward greedy methods. Our convergence analysis also holds here.

2.5.1 Proof of convergence

A trivial upper bound on the number of iterations of NOMP is the number of the observations N. This is a direct result of solving the least squares at Step 10 of the algorithm. After N iterations, X is a square full-rank matrix (no frequency is detected twice), hence the residue is zero for $(N+1)^{th}$ iteration and the algorithm terminates.

The following theorem states another upper bound on the number of iterations, obtained by characterizing the amount by which the residual energy decreases when adding a new frequency to the set of estimated sinusoids.

Theorem 1 *The reduction of residual energy due to one iteration (adding a new sinusoid) is at least τ. Consequently, $\min\{N, \|\mathbf{y}\|^2/\tau\}$ is an upper bound on the number of iterations of the algorithm.*

Proof: *The residue at m^{th} iteration of the algorithm is given by $\mathbf{y}_r(\mathcal{P}_m) = \mathbf{y} - \sum_{l=1}^{l=m} g_l \mathbf{x}(\omega_l)$. The energy of the residue in each iteration of the algorithm satisfies the following,*

$$\begin{aligned}
\|\mathbf{y}_r(\mathcal{P}_{m-1})\|^2 &\overset{(a)}{=} \|\mathbf{y}_r(\mathcal{P}'_m)\|^2 + G_{\mathbf{y}_r(\mathcal{P}_{m-1})}(\hat{\omega}) \\
&\overset{(b)}{\geq} \|\mathbf{y}_r(\mathcal{P}'''_m)\|^2 + G_{\mathbf{y}_r(\mathcal{P}_{m-1})}(\hat{\omega}) \\
&\overset{(c)}{\geq} \|\mathbf{y}_r(\mathcal{P}_m)\|^2 + G_{\mathbf{y}_r(\mathcal{P}_{m-1})}(\hat{\omega}) \qquad (2.16) \\
&\overset{(d)}{\geq} \|\mathbf{y}_r(\mathcal{P}_m)\|^2 + \tau. \qquad (2.17)
\end{aligned}$$

*where (a) follows from Step 5 of the algorithm where we project $\mathbf{y}_r(\mathcal{P}_{m-1})$ orthogonal to the subspace spanned by $\mathbf{x}(\hat{\omega})$ to get $\mathbf{y}_r(\mathcal{P}'_m)$. Inequalities in (b) follow from (**RAC**) checks performed whenever the single frequency refinement algorithm is invoked and (c) from the fact that least squares* Update *can only lead to a decrease in energy of the residual signal. (d) is a direct consequence of the stopping criteria of the algorithm.*

Inequality (2.17) shows that the reduction of the residual energy due to detecting a new sinusoids is always greater than τ. This result bounds the number of iterations of the algorithm from above by $\|\mathbf{y}\|^2/\tau$, proving convergence. Combining this observation with the trivial upper-bound N completes the proof.

2.5.2 Rate of convergence

We first bound the maximum of the GLRT cost function $G_{\mathbf{y}}(\omega)$ over the continuum of frequencies, in terms of that over the oversampled grid. To this end, we borrow ideas used in [19] (Appendix C) for proving a similar property for the dual atomic norm. We briefly introduce the notion of atomic norm (also known as dictionary norm) [60] and specialize to the line spectral estimation problem.

The atomic set of unit norm sinusoids is given by $\mathcal{A} = \{e^{j\phi}\mathbf{x}(\omega) : \phi, \omega \in [0, 2\pi)\}$. The atomic norm for $\mathbf{y}$ is defined by

$$\|\mathbf{y}\|_{\mathcal{A}} \triangleq \inf \{t > 0 : \mathbf{y} \in t \ \mathrm{conv}(\mathcal{A})\}. \tag{2.18}$$

where $\mathrm{conv}(\mathcal{A})$ denotes the convex hull of the points in $\mathcal{A}$. Since the centroid of the $\mathrm{conv}(\mathcal{A})$ is at the origin, the atomic norm can be rewritten as [71, 72]

$$\|\mathbf{y}\|_{\mathcal{A}} \triangleq \inf \left\{ \sum_l |g_l| : \mathbf{y} = \sum_l g_l \mathbf{x}(\omega_l),\ \mathbf{x}(\omega_l) \in \mathcal{A} \right\}. \tag{2.19}$$

Note that this is not the ℓ_1 norm of $\mathbf{y}$, but the ℓ_1 norm on the coefficients of the representation of $\mathbf{y}$ by elements of $\mathcal{A}$. The atomic norm is typically small when its argument has a good sparse approximation [13]. The dual norm of $\|\cdot\|_{\mathcal{A}}$ is defined by

$$\|\mathbf{y}\|_{\mathcal{A}}^* = \sup_{\mathbf{a} \in \mathcal{A}} \Re\{\langle \mathbf{a}, \mathbf{y} \rangle\}. \tag{2.20}$$

It is easy to see that

$$\begin{aligned} \|\mathbf{y}\|_{\mathcal{A}}^{*} &= \sup_{\omega\in[0,2\pi)} \sup_{\phi\in[0,2\pi)} \Re\{e^{i\phi}\langle \mathbf{x}(\omega), \mathbf{y}\rangle\} \\ &= \sup_{\omega\in[0,2\pi)} |\langle \mathbf{x}(\omega), \mathbf{y}\rangle| \\ &= \sup_{\omega\in[0,2\pi)} \sqrt{G_{\mathbf{y}}(\omega)}. \end{aligned} \tag{2.21}$$

Directly borrowing from [19] gives us the following Theorem.

Theorem 2 *[19] Maximizing the GLRT cost function (for the dictionary of unit norm sinusoids) over $[0, 2\pi)$ is consistent with that over the oversampled grid Ω with oversampling factor γ. That is, we have*

$$\max_{\omega\in\Omega} \sqrt{G_{\mathbf{y}}(\omega)} \leq \sup_{\omega\in[0,2\pi)} \sqrt{G_{\mathbf{y}}(\omega)} \tag{2.22}$$

$$\leq \left(1 - \frac{2\pi}{\gamma}\right)^{-1} \max_{\omega\in\Omega} \sqrt{G_{\mathbf{y}}(\omega)}. \tag{2.23}$$

See ([19], Appendix C) for a proof.

We need the following lemma to prove Theorem 3, which provides a pessimistic characterization of the convergence rate.

Lemma 1 *Assume $\{a_n\}_{n\geq 0}$ is a decreasing sequence of nonnegative numbers such that $a_0 \leq U$ and*

$$a_n \leq a_{n-1}\left(1 - \frac{a_{n-1}}{U}\right), \quad \forall n > 0,$$

then we have $a_n \leq \frac{U}{n+1}$ for all $n \geq 0$.

The proof is by induction [60]. Suppose $a_{n-1} \leq \frac{U}{n}$. Either $a_{n-1} \leq \frac{U}{n+1}$, in which case

$a_n \leq \frac{U}{n+1}$, *or* $a_{n-1} \geq \frac{U}{n+1}$, *in which case*

$$\begin{aligned} a_n &\leq a_{n-1}\left(1-\frac{a_{n-1}}{U}\right) \\ &\leq \frac{U}{n}\left(1-\frac{\frac{U}{n+1}}{U}\right) = \frac{U}{n+1}. \end{aligned}$$

Hence, $a_n \leq \frac{U}{n+1}$ *for all* $n \geq 0$.

Theorem 3 *For all* $\mathbf{y}$ *such that* $\|\mathbf{y}\|_{\mathcal{A}} < \infty$, *the residual energy of NOMP at the* m^{th} *iteration satisfies*

$$\|\mathbf{y}_r(\mathcal{P}_m)\| \leq (m+1)^{-1/2}\left(1-\frac{2\pi}{\gamma}\right)^{-1}\|\mathbf{y}\|_{\mathcal{A}}. \tag{2.24}$$

Proof: *From (2.16), we have*

$$\|\mathbf{y}_r(\mathcal{P}_m)\|^2 \leq \|\mathbf{y}_r(\mathcal{P}_{m-1})\|^2 - G_{\mathbf{y}_r(\mathcal{P}_{m-1})}(\hat{\omega}). \tag{2.25}$$

$\mathbf{y}_r(\mathcal{P}_{m-1})$ *is the result of projecting* $\mathbf{y}$ *orthogonal to the subspace spanned by* $\mathcal{P}_{m-1}$, *therefore*

$$\begin{aligned} \|\mathbf{y}_r(\mathcal{P}_{m-1})\|^2 &= \langle \mathbf{y}_r(\mathcal{P}_{m-1}), \mathbf{y}\rangle \\ &\overset{(a)}{\leq} \|\mathbf{y}\|_{\mathcal{A}} \|\mathbf{y}_r(\mathcal{P}_{m-1})\|_{\mathcal{A}}^* \\ &= \|\mathbf{y}\|_{\mathcal{A}} \sup_{\omega\in[0,2\pi)} \sqrt{G_{\mathbf{y}_r(\mathcal{P}_{m-1})}(\omega)} \\ &\overset{(b)}{\leq} \|\mathbf{y}\|_{\mathcal{A}}\left(1-\frac{2\pi}{\gamma}\right)^{-1} \max_{\omega\in\Omega} \sqrt{G_{\mathbf{y}_r(\mathcal{P}_{m-1})}(\omega)}, \end{aligned}$$

where (a) *follows by Hölder's inequality [73], and* (b) *is by Theorem 2.*

Let $\eta \triangleq \|\mathbf{y}\|_{\mathcal{A}} \left(1 - \frac{2\pi}{\gamma}\right)^{-1}$. *From Step 5 of the algorithm we have*

$$\hat{\omega} = \arg\max_{\omega \in \Omega} \sqrt{G_{\mathbf{y}_r(\mathcal{P}_{m-1})}(\omega)},$$

hence,

$$\|\mathbf{y}_r(\mathcal{P}_{m-1})\|^2 \leq \eta \sqrt{G_{\mathbf{y}_r(\mathcal{P}_{m-1})}(\hat{\omega})}. \tag{2.26}$$

Combining (2.25) and (2.26), gives

$$\|\mathbf{y}_r(\mathcal{P}_m)\|^2 \leq \|\mathbf{y}_r(\mathcal{P}_{m-1})\|^2 \left(1 - \eta^{-2} \|\mathbf{y}_r(\mathcal{P}_{m-1})\|^2\right). \tag{2.27}$$

Using Lemma 1 and the fact that

$$\|\mathbf{y}_r(\mathcal{P}_0)\|^2 = \|\mathbf{y}\|^2 \leq \|\mathbf{y}\|_{\mathcal{A}}^2 \leq \eta^2, \tag{2.28}$$

we have

$$\|\mathbf{y}_r(\mathcal{P}_m)\|^2 \leq \frac{\eta^2}{m+1}. \tag{2.29}$$

In other words,

$$\|\mathbf{y}_r(\mathcal{P}_m)\| \leq (m+1)^{-1/2} \left(1 - \frac{2\pi}{\gamma}\right)^{-1} \|\mathbf{y}\|_{\mathcal{A}}. \tag{2.30}$$

This proves the Theorem.

For the simulations we set $\gamma = 4$, but it is worth mentioning that, since we employ FFTs for detection over the oversampled grid, increasing γ has marginal effect on the runtime of the algorithm. In fact setting $\gamma = 20$ leads to only about 5% increase in runtime. If we compare (2.30) to the rate of convergence of OMP over the continuum

given by [60],

$$\|\mathbf{y}_{\mathrm{r}}(\mathcal{P}_m)\|^2 \leq (m+1)^{-1} \|\mathbf{y}\|_{\mathcal{A}}^2 , \tag{2.31}$$

we see that by choosing γ large enough, the bound on the convergence rate specified in (2.30) approaches that of OMP over the continuum. Note that, in the derivation of (2.30) we have not considered the effect of the refinement steps, but imposing `RAC` ensures that refinements can only help speed up convergence.

2.5.3 Empirical rate of convergence

We use numerical simulations to highlight the convergence benefits of the refinement steps in NOMP compared to the bound specified in (2.30). We plot the mean residual energy (averaged over 1000 runs) as a function of the number of iterations in a noiseless setting. We set $K = 16$, and $\Delta\omega_{\min} = 2.5 \times \Delta_{\mathrm{dft}}$. Figure 2.5 shows that light oversampling ($\gamma = 4$) followed by Single Refinement step (NOMP–) leads to a better convergence rate than having a large oversampling factor ($\gamma = 20$) and no refinements (DOMP). On the other hand, NOMP enjoys an extremely fast convergence rate due to the Cyclic Refinement step. In fact we see that for the setting where $\Delta\omega_{\min}$ is fairly large, the residual energy essentially drops to machine precision after 16 iterations, which equals the number of sinusoids in the mixture.

2.6 Simulation Results

Our performance measure is the mean squared error (MSE) of frequency estimation, and we compare the performance of NOMP against a number of benchmarks in various settings.

Benchmarks: The MUSIC algorithm is implemented using a modified version of the

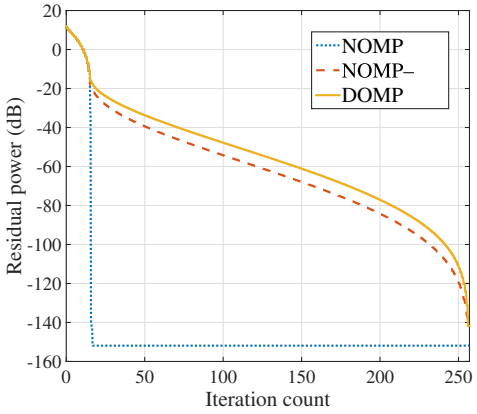


Figure 2.5: Convergence rates at noiseless case

MATLAB routine `rootmusic`. The modifications are two-fold: (1) we use Minimum Description Length (MDL) criterion [74] for estimating the number of sinusoids in the mixture; (2) MUSIC operates by constructing an estimate of the autocorrelation matrix of the observed data vector. To this end, we use a sliding window of size W, to generate multiple snapshots from the observation vector $\mathbf{y} \in \mathbb{C}^N$. The choice of W has a significant impact on the performance of the algorithm, both in terms of estimation accuracy and run time: W too large leads to an inaccurate estimate of the autocorrelation matrix and the signal subspace, whereas too small a value effectively reduces the size of the observation window in time, degrading frequency estimation accuracy. We have found (empirically) that setting $W = 96$ results in the best estimation accuracy for the nominal settings in our simulations.

For sparse convex optimization, we consider Atomic norm Soft Thresholding (AST) [19, 75]. We use the alternating direction method of multipliers (ADMM) [76] to implement AST, as suggested in [19]. The updates in ADMM are iterative (typically $100 \sim 200$ iterations for each simulation run), and each iteration includes an eigenvalue thresholding step for an $(N+1) \times (N+1)$ matrix. This $\mathcal{O}(N^3)$ step dominates the computational cost of the ADMM method, and becomes very expensive for large N.

The authors in [19] suggest solving Lasso as an alternative to the semi-definite program induced by AST. Lasso is solved on an oversampled frequency grid, using the highly optimized $\ell_2 - \ell_1$ software package SpaRSA [77]. We set the tolerance parameter to be 10^{-3} (other than this, we use the default parameters): smaller values of the tolerance parameter (e.g. 10^{-4}) increase runtime significantly, while providing marginal performance improvement. The regularization parameter in AST and Lasso formulation, suggested in [19], is set to

$$\text{reg} = \sigma \left(1 + \frac{1}{\log N}\right) \sqrt{\log N + \log(4\pi \log N)}.$$

The oversampling factor for the Lasso solver is set to 10 in our simulations. While increasing the oversampling factor improves Lasso performance, as mentioned in [19], for an oversampled grid, the frequencies estimated by Lasso cluster around the true frequencies. In order to avoid drastically overestimating model order, we implement a simple *clustering* scheme. In our simulations, we group frequencies into the largest number of clusters possible so that the frequency separation between any two sinusoids in different clusters is no smaller than $0.25 \times \Delta_{\text{dft}}$. After clustering, we update the gains of the cluster centers by solving the least squares problem $\text{minimize}_{\{g_l\}} \|\mathbf{y} - \sum_l g_l \mathbf{x}(\omega_l)\|^2$.

Newtonized Lasso (NLasso): We also compare the results of NOMP algorithm with an extension of the Lasso formulation. In this scheme, we first apply the Lasso solver to identify the frequencies over the highly oversampled grid. Then we run the Cyclic Refinement step of the NOMP algorithm in order to refine the estimated frequencies, in order to prevent error floors caused by the off-grid effect. The parameters of Lasso are unchanged and the number of refinement steps is set to 5. Our simulations show that, for well-separated frequencies, the refinement step significantly improves estimation accuracy while incurring a small increase in runtime compared to Lasso. However, the benefit of refinement for Lasso diminishes as we increase the difficulty of the estimation

problem (small $\Delta\omega_{\min}$).

Simulation set-up: We consider a mixture of $K = 16$ sinusoids of length $N = 256$. We perform 300 simulation runs for each of the four scenarios characterized by $\Delta\omega_{\min}$ and SNR values. The settings for different scenarios are summarized in Table 2.1.

Scenarios	SNR (dB)	$\Delta\omega_{\min}/\Delta_{\text{dft}}$
1	SNR_{nom}	2.5
2	SNR_{nom}	0.5
3	Uniform$[15, 35]$	2.5
4	Uniform$[15, 35]$	0.5

Table 2.1: Settings of different scenarios.

In Scenarios 1 and 2, the nominal SNR for each sinusoid is set as $\text{SNR}_{\text{nom}} = 25$ dB, whereas for Scenarios 3 and 4, the SNR values are chosen uniformly from $[15, 35]$ dB, with mean equal to the nominal SNR of 25 dB. In each simulation run, the gain magnitudes are set to $|g_l| = \sigma\sqrt{\text{SNR}_l}$, while the phases $\{\angle g_l\}$ are chosen uniformly from $[0, 2\pi)$. The frequencies are chosen uniformly at random from $[0, 2\pi)^K$ while respecting the minimum separation constraints specified by $\Delta\omega_{\min}$ (if the minimum separation criterion is not met, we sample again from $[0, 2\pi)^K$). We plot the Complementary Cumulative Distribution Function (CCDF) of the squared frequency estimation error for all algorithms, along with the CRB (also a random variable, since it differs across realizations), and also compare against the DFT spacing, which is the resolution provided by coarse peak picking. See Appendix II for a quick overview of Cramér Rao Bound. The parameters of NOMP algorithm are set according to Table 2.2 for different scenarios.

NOMP	Scenario 1	Scenario 2	Scenario 3	Scenario 4
R_c	1	3	1	3
R_s	1	1	1	1
γ	4	4	4	4

Table 2.2: NOMP parameters at different scenarios.

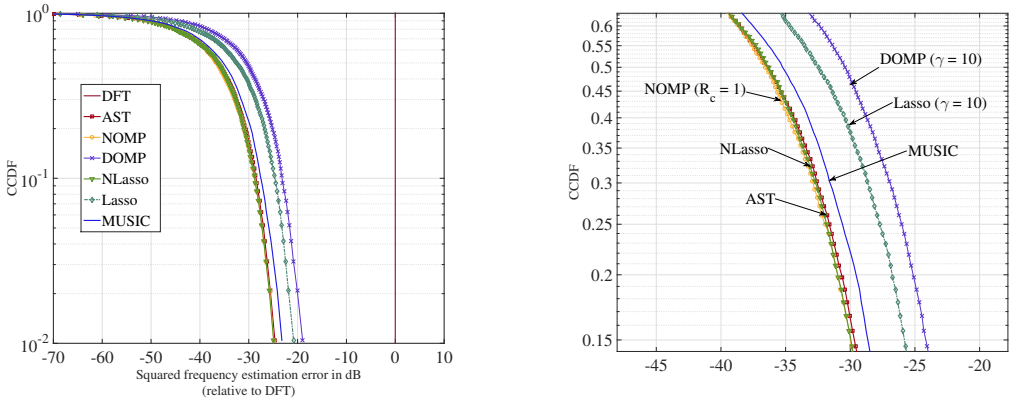


Figure 2.6: CCDF of the frequency MSE for Scenario 1 (figure on the right is zoomed in).

2.6.1 Frequency estimation accuracy

Distribution of error

Let us first examine the CCDF of squared frequency estimation error in each scenario. Figure 2.6 shows that NOMP, AST and NLasso lead to very similar error distributions in Scenario 1, while outperforming the other methods. Unlike Lasso and DOMP, which suffer from the off-grid effect, MUSIC picks frequencies over the continuum, and achieves better estimation accuracy. Another observation is that if the frequencies are well separated (as in Scenario 1), then adding a refinement stage at the output of Lasso leads to a significant improvement in estimation accuracy. As we move to more difficult Scenarios, however, the performance of NLasso degrades compared to NOMP and AST as shown in Figure 2.7 for Scenario 2. The reason is that the refinement stage of NLasso is able to improve the estimation accuracy only when the initial estimates provided by Lasso are close to the true frequencies. When Lasso fails in providing good initial estimates, there is no benefit in locally refining the frequencies.

Figure 2.8 shows the distribution of error in Scenario 3. We see that the overall gap in the performance of different algorithms is reduced compared to the first two scenarios,

with AST, NOMP and NLasso still achieving the highest estimation accuracy. In Scenario 4, NOMP achieves superior performance compared to all of the other methods, as shown in Figure 2.9.

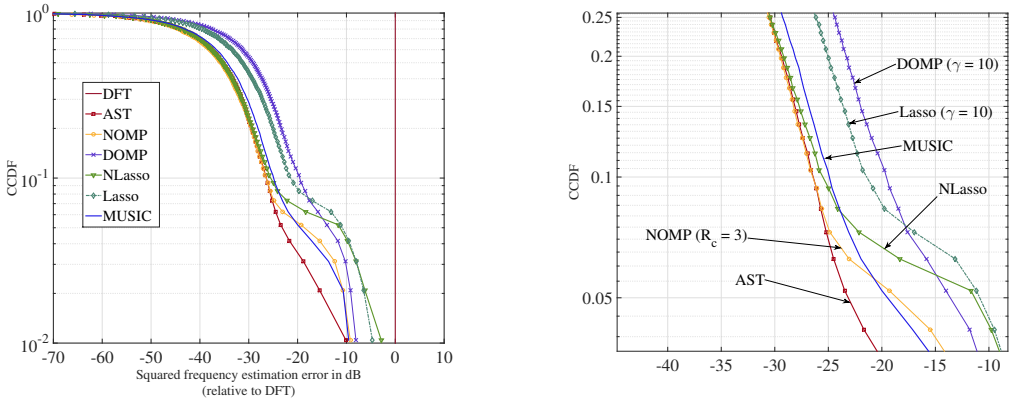


Figure 2.7: CCDF of the frequency MSE for Scenario 2 (figure on the right is zoomed in).

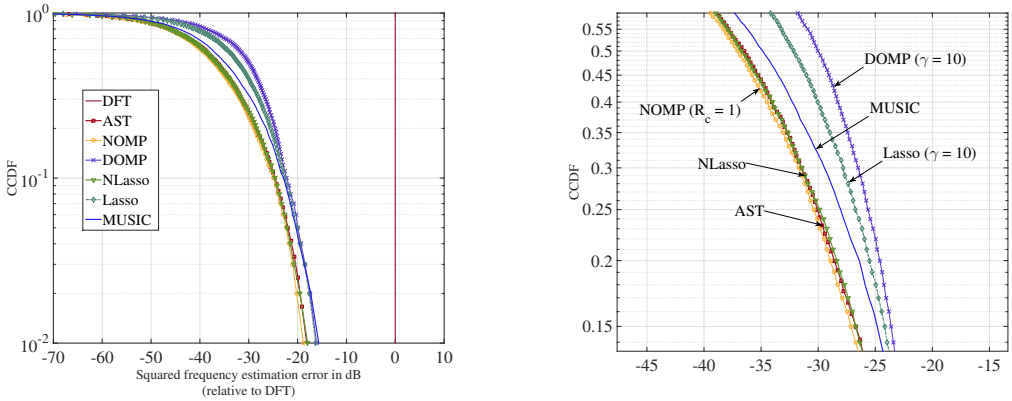


Figure 2.8: CCDF of the frequency MSE for Scenario 3 (figure on the right is zoomed in).

Mean squared error

Here we examine the performance of different algorithms in terms of frequency estimation accuracy by looking at the normalized Mean Squared Error (MSE), defined by $\mathbb{E}[(\omega_{true} - \omega_{est})^2]/\Delta_{\text{dft}}^2$, in different scenarios. In Scenarios 1 and 3, where frequencies are

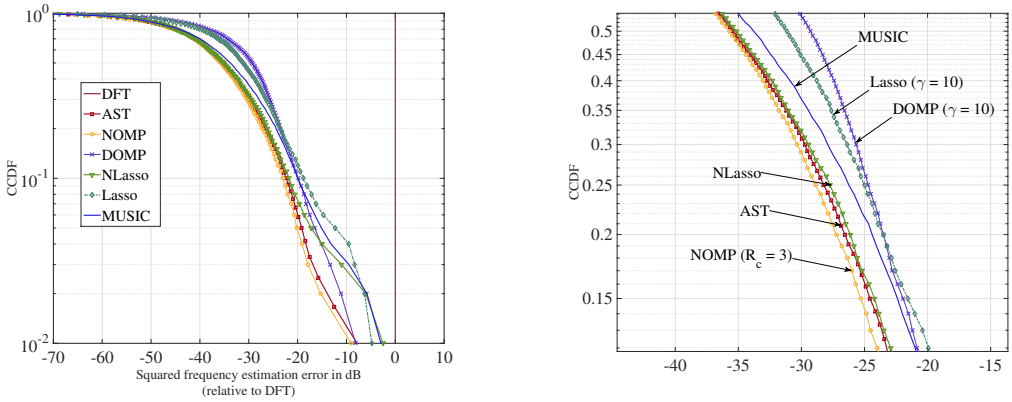


Figure 2.9: CCDF of the frequency MSE for Scenario 4 (figure on the right is zoomed in).

well-separated, one hopes to get to the estimation accuracy of a single sinusoid (as if the other sinusoids did not exist). We therefore use the CRB and ZZB corresponding to a *single sinusoid*, computed by (2.10) and (2.11), respectively, as measures of optimality. In Scenarios 2 and 4, where frequencies can get close to one another, we compute the CRB empirically for each realization of the problem, and employ the mean CRB as a measure of optimality.

Figure 2.10 shows the MSE of frequency estimation in Scenarios 1 and 2, for $\mathrm{SNR}_{\mathrm{nom}}$ taking values from 13 dB to 35 dB. In Scenario 1, if the nominal SNR is high enough, AST, NOMP, and NLasso all achieve the CRB. As we decrease the nominal SNR, all of the algorithms exhibit threshold behavior, well predicted by the ZZB threshold. The threshold SNR of AST is lower than that of other methods, showing its noise resilience. MUSIC does not achieve the CRB, but closely follows the bound for all SNR values in this scenario. DOMP and Lasso on the other hand, reach performance floors. We examine this floor more closely in Section 2.6.3 to see whether it is a fundamental algorithmic limitation, or happens due to the off-grid effect. Figure 2.10-b corresponds to Scenario 2, where the separation between frequencies can be very small. We see that MUSIC is the only algorithm that benefits from increasing the nominal SNR. This goes back to

the asymptotic optimality of MUSIC: for $SNR \to \infty$ and $K \ll N$, MUSIC is able to precisely determine the frequencies in the mixture, regardless of the separation between them [34].

In Scenarios 3 and 4, the SNRs are drawn independently and uniformly at random from the interval $[15, 35]$ dB. In order to evaluate the frequency MSE in these scenarios, we fix the SNR of one of the sinusoids in the mixture at a given value, while letting the other $(K-1)$ SNRs to be realized randomly. Figure 2.11-a shows the frequency MSE curves corresponding to Scenario 3. As we expected from error distribution plots in Figure 2.8, the gap in the performance of different algorithms has decreased compared to the first scenario. Figure 2.11-b corresponds to Scenario 4, in which NOMP outperforms all of the other algorithms, and tightly follows the CRB. This indicates that NOMP is highly successful in exploiting the disparity in SNRs across sinusoids in the mixture, in order to estimate closely-spaced frequencies. AST achieves the best performance at very low SNRs; however, its MSE curve stays bounded away from the CRB, with an expanding gap as we increase SNR.

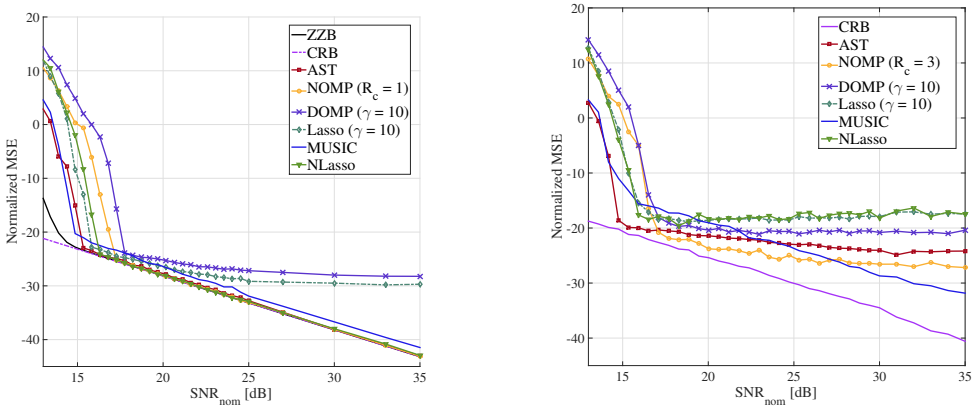


Figure 2.10: Normalized frequency MSE for (left) Scenario 1, and (right) Scenario 2.

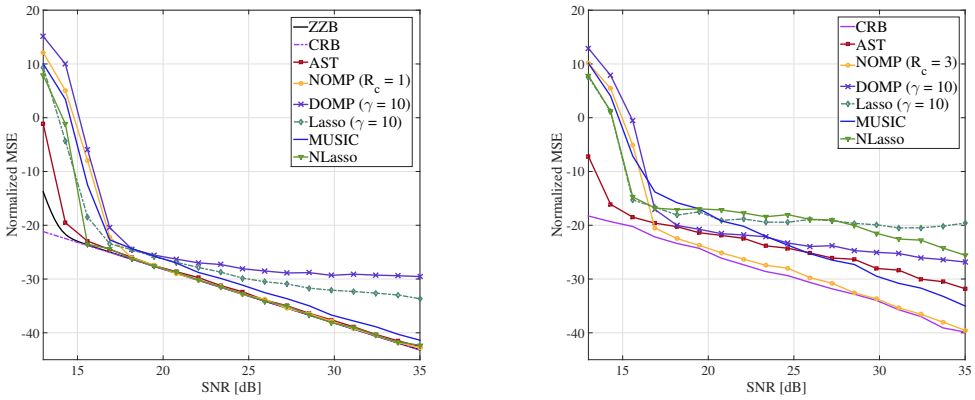


Figure 2.11: Normalized frequency MSE for (left) Scenario 3, and (right) Scenario 4.

Number of cycles of Newton refinement (R_c)

We have seen that NOMP is able to achieve the CRB in Scenarios 1 and 3, with one cycle of refining the sinusoids in each iteration, i.e., $R_c = 1$. In this subsection, we want to highlight the effect of increasing R_c in improving the frequency estimation accuracy of NOMP in Scenarios 2 and 4 where $\Delta\omega_{\min} = 0.5 \times \Delta_{\text{dft}}$. Figure 2.12 shows the frequency MSE of NOMP for $R_c \in \{1, 3, 5\}$. We see that NOMP enjoys the benefits of having more rounds of refinement, but exhibits diminishing return as R_c increases. In particular, increasing R_c beyond 3 cycles gives marginal improvement in estimation accuracy.

2.6.2 Computation time

Table 2.3 summarizes the time needed for running 300 simulations for each of the algorithms in different scenarios. We see that DOMP is extremely fast at the expense of estimation accuracy. NOMP is faster than all of the other methods in all four Scenarios, while achieving remarkable frequency estimation accuracy. As we discussed in Section 2.2.3, the Cyclic Refinement step has the complexity $\mathcal{O}(R_c R_s K^2 N)$, and dominates

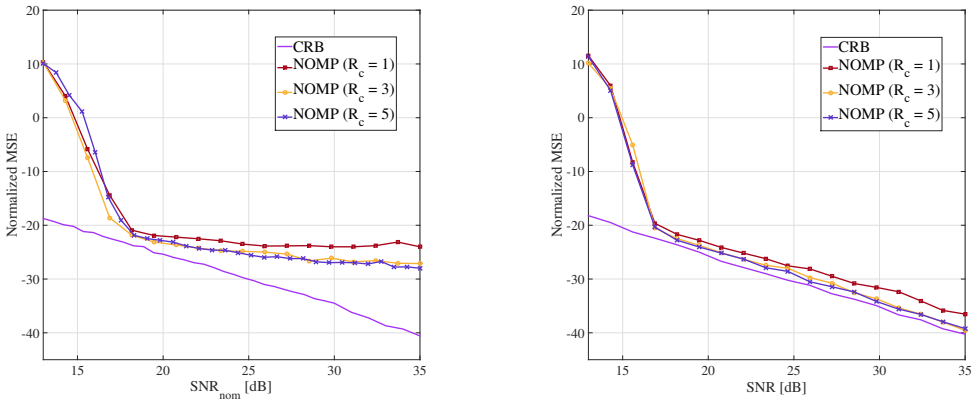


Figure 2.12: Performance improvement of NOMP with increasing the number of cyclic refinements in (left) Scenario 2, and (right) Scenario 4.

the computational cost of NOMP. Table 2.4 shows the time needed for 300 simulation runs of NOMP for different values of R_c. Note that as we increase the difficulty of the estimation scenario, AST, Lasso and NLasso tend to take more time, while MUSIC and NOMP are unaffected.

Time [sec]	NOMP	AST	NLasso	Lasso	DOMP	MUSIC
Scenario 1	6.92	1.09e3	29.68	26.63	2.66	19.83
Scenario 2	14.26	1.02e3	29.63	27.62	2.81	20.15
Scenario 3	6.88	1.12e3	34.32	32.06	2.79	20.07
Scenario 4	14.19	1.18e3	36.37	33.71	2.80	19.69

Table 2.3: Time [sec] for 300 runs of each algorithm. Parameters of NOMP are set by Table 2.2

Time [sec]	$R_c = 1$	$R_c = 3$	$R_c = 5$
Scenario 1	6.92	14.24	21.67
Scenario 2	7.01	14.26	21.71
Scenario 3	6.88	14.21	21.61
Scenario 4	6.96	14.19	21.50

Table 2.4: Time [sec] for 300 runs of NOMP algorithm, for different numbers of cyclic refinements R_c.

2.6.3 Asymptotic regime

It is interesting to see the effect of increasing the oversampling factor for Lasso and DOMP. Figure 2.13 corresponds to Scenario 1, with the change being that the oversampling factor for these two algorithms is increased. We observe an improvement in the performance of both algorithms in terms of estimation accuracy, with that for Lasso being especially significant. Comparing these results to those in Figure 2.10-a, we see that the MSE performance of DOMP is marginally improved by increasing γ from 10 to 100. This performance plateau shows a fundamental *algorithmic limitation* of DOMP, and highlights the critical role of cyclic Newton refinements in NOMP. In other words, the performance limitation of DOMP is not just due to the off-grid error, but also a consequence of making "hard-decisions" at each iteration. The computational complexity of DOMP is insensitive to oversampling factor. For example, computation time increases from 3.25 to about 7.96 seconds as we go from $\gamma = 20$ to $\gamma = 100$.

In theory, Lasso is only limited by the oversampling factor γ, and as $\gamma \to \infty$, the result of Lasso converges to that of AST, as a consequence of the convergence of the corresponding atomic norms [19]. In practice, however, the performance of Lasso solved on a moderate size grid might be far from that of AST. Figure 2.13 shows that, at large enough oversampling factors, Lasso approaches the CRB, but the computational cost becomes prohibitive. For example, the computation time of Lasso for 300 runs, increases from 72.16 seconds to 191.22 seconds as we go from $\gamma = 20$ to $\gamma = 50$.

2.6.4 Model order estimation

Estimating the model order K (the true number of non-zero atoms in the mixture) has significant importance in sparse approximation. Here we examine the Cumulative Distribution Function (CDF) of the estimated model order by different algorithms. As

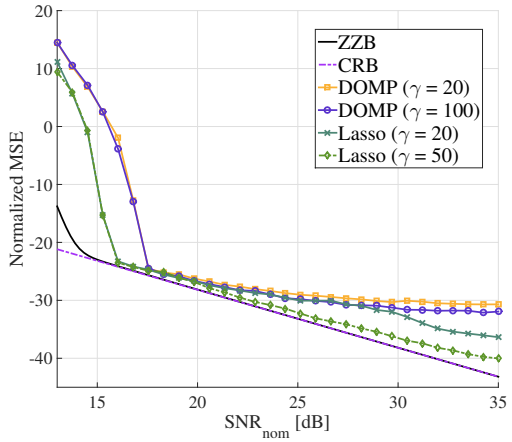


Figure 2.13: Frequency MSE for Scenario 1 and highly over-sampled grid for Lasso and DOMP.

shown in Figure 2.14, both AST and MUSIC perform well, with MUSIC performance slightly degrading in scenarios with small $\Delta\omega_{\min}$. Note that the mean of the distributions are very close to the truth (small bias), and they have a small spread around the mean (small variance). Figure 2.15 shows the model order estimates for NOMP using both CFAR and BIC-based stopping criteria. We see that both criteria are very accurate in all four scenarios.

On the other hand, Lasso and DOMP perform poorly in estimating the model order, especially in scenarios with small $\Delta\omega_{\min}$. As shown in Figure 2.16, DOMP tends to overestimate the model order when some of the frequencies are placed too closely. This is again the result of a fundamental algorithmic limitation. DOMP does not allow for correcting errors that have happened in the previous iterations; instead, it tries to explain the residual energy by overestimating the number of non-zero atoms.

In Scenario 2, Lasso tends to underestimate the model order. The main reason is that for two closely spaced frequencies, Lasso generates two overlapping clusters of estimated frequencies, which are later replaced by a single frequency by our *clustering* algorithm. On the other hand, if we do not employ clustering, Lasso significantly overestimates the

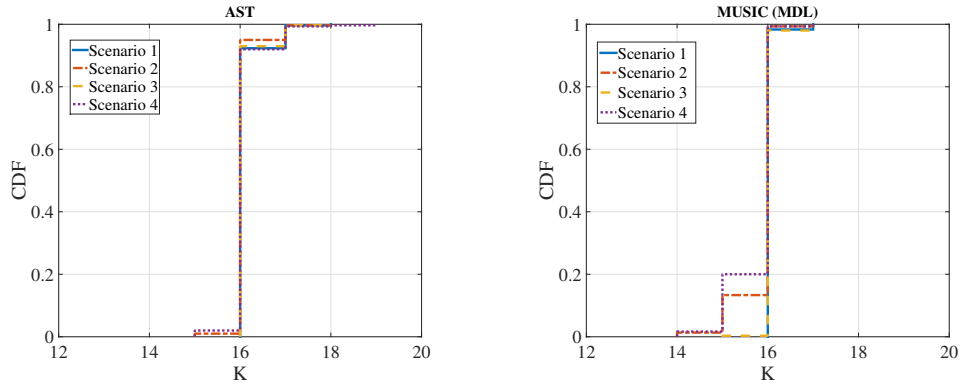


Figure 2.14: CDF of the estimates of the model order for (left) AST, and (right) MUSIC.

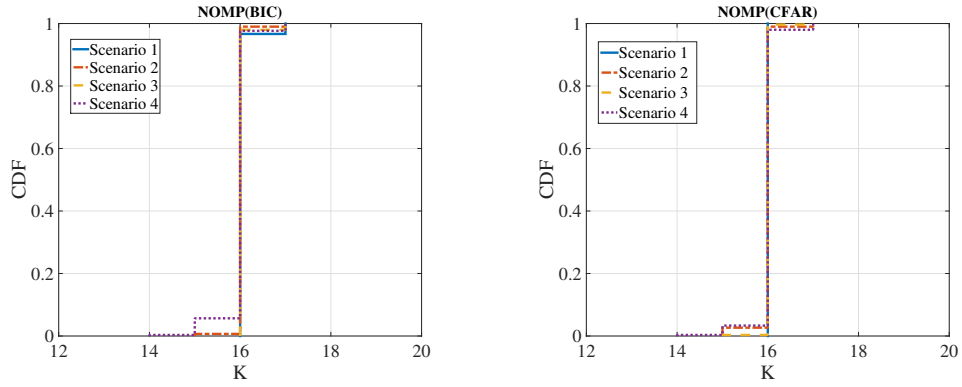


Figure 2.15: CDF of the estimates of the model order for NOMP using (left) BIC, and (right) CFAR criteria.

model order.

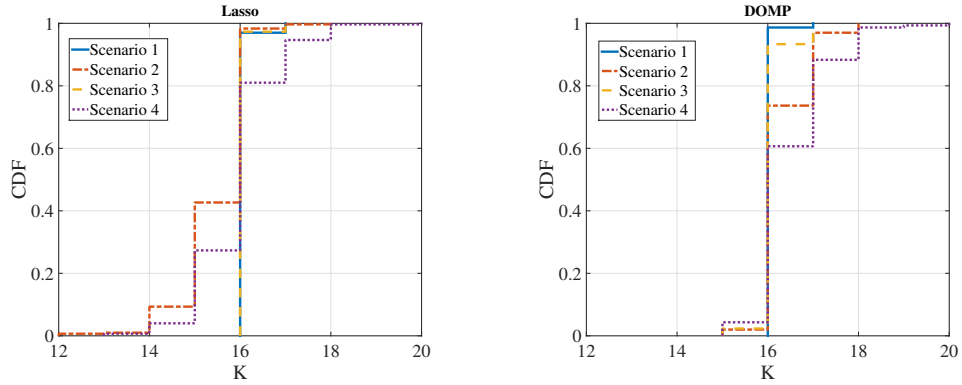


Figure 2.16: CDF of the estimates of the model order for (left) Lasso, and (right) DOMP with $\gamma = 20$.

2.7 Extensions of the Algorithm

In this section we point out an immediate extension of NOMP algorithm. Specifically, we can replace the manifold of sinusoids $\{\mathbf{x}(\omega) \ : \ \omega \in [0, 2\pi)\}$ by

$$\{A\mathbf{x}(\omega)/\left\|A\mathbf{x}(\omega)\right\| \ : \ \omega \in [0, 2\pi)\},$$

where $A \in \mathbb{C}^{M\times N}$ is a known measurement matrix. This is motivated by the measurement setup $\mathbf{y} = \sum_{l=1}^{K} g_l A\mathbf{x}(\omega_l)$.

Compressive measurements

We consider a compressive measurement model in which the number of measurements $M \ll N$. As in the bulk of literature on compressive sensing, we assume that the elements of A are chosen i.i.d from appropriate zero-mean distributions (with variance conveniently scaled to $1/N$ such as Uniform$\{\pm 1/\sqrt{N}\}$, Uniform$\{\pm 1/\sqrt{N}, \pm j/\sqrt{N}\}$,

$\mathcal{N}(0, 1/N)$, etc.,) so that certain concentration results hold. It has been shown in [64] that when A satisfies certain isometry conditions (related to the estimation problem at hand), the CRB and ZZB are approximately preserved for compressive estimation, except for an SNR degradation of M/N due to randomly projecting down the signal to a smaller space. The number of compressive measurements needed to estimate continuous frequencies scales as $M = \mathcal{O}(K \log N)$ [64]. In order to get concrete numerical intuition, we set $M = N/4 = 64$ with the elements of A chosen uniformly and independently at random from $\{\pm 1/\sqrt{N}, \pm j/\sqrt{N}\}$ and run NOMP with atoms $\{\mathbf{s}(\omega) = A\mathbf{x}(\omega)/\left\|A\mathbf{x}(\omega)\right\| : \omega \in [0, 2\pi)\}$. We consider Scenario 1 with $R_c = 3$, with number of sinusoids set to $K = 13$ and $K = 16$. Our algorithm approaches the CRB in the setting where $K = 13$, whereas we incur large estimation errors when $K = 16$; see Figure 2.17. The large estimation errors for $K = 16$ occur because the compressive measurement matrix A fails to preserve the structure of the estimation problem: $M = 64$ compressive measurements is too few for $K = 16$ sinusoids.

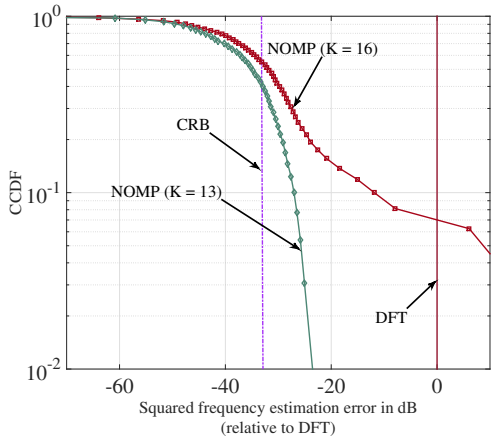


Figure 2.17: CCDF of the frequency MSE for Scenario 1 with Compressive measurements.

Chapter 3

Sparsity Based Imaging

3.1 Introduction

In this chapter, we develop a theoretical framework for an array of monostatic elements for short-range mm-wave radar imaging, which has significant potential for emerging applications such as handheld-based gesture recognition and vehicular situational awareness. We show that we must account for form factor and complexity constraints in a manner that is quite different from that of conventional radar, and introduce *new models and algorithms* validated by experimental results. Our main results include (a) identification of the number of degrees of freedom as a function of scene and array geometry, (b) demonstrating that grating lobes seen for the conventional point scatterer model when we employ sparse arrays can be eliminated by using spatial aggregation, using patch-based scatterer models as our dictionary, (c) optimization of patch size based on estimation-theoretic bounds, (d) sparse reconstruction using NOMP algorithm (described in Chapter 2 for the problem of frequency estimation), benchmarked against a standard ℓ_1-regularized convex optimization framework.

3.1.1 Related work

The deterioration of imaging performance due to grating lobes caused by undersampling is well known [78, 79]. Prior approaches for grating lobe suppression include the use of shaped waveforms and aperture diversity [78], employing a multistatic array (synchronization) [80], frequency diversity (wideband) [81], and incorporating Doppler information from moving targets [82]. However, to our knowledge, there is little prior work done on improving scene *modeling* to combat grating lobes. In our previous work [28, 83], we introduced some of the concepts underlying this chapter, and provided some experimental results. Here, we take the opportunity of elaborating on theoretical and algorithmic aspects.

The focus of this chapter is to extract the information from a *static scene* by reconstructing the spatial configuration of the objects and creating an image of the scene. This is quite different from the scenario in which the desired information resides in the temporal changes of the received signal, e.g., for recognizing *motion gestures* of a human hand [4], or a pen [5]. Our spatial aggregation and sparse reconstruction approach may be viewed as falling within the general framework of synthesis-based sparse signal representation, which has been an active area of research for the past decade or so [84, 85, 86]. One of the main advantages of this approach is that it explicitly incorporates prior information regarding the nature of the aperture (e.g., sparsity level and the geometry of the array), and the nature of the scene (e.g., shape/size/type of targets in the scene) for dictionary construction and image formation. An implicit assumption behind our patch-based primitive is that the scene reflectivity is *lowpass* (i.e, it varies slowly in space). Finally, we investigate the "sparse reconstruction" framework [13, 87, 88] that is based on finding the maximally sparse representation of the scene in the dictionary of spatially extended objects. To this end, we employ an ℓ_1-regularized convex optimization method

[89], as well as a variant of the NOMP algorithm [32] described in Chapter 2. While the NOMP algorithm was shown to represent the state of the art in terms of empirical performance for a classical frequency estimation problem, more detailed comparison with competing algorithms for alternative models remains an open issue beyond our current scope.

The rest of the chapter is organized as follows. In Section 3.2, we explain the grating lobe problem associated with sparse monostatic imaging arrays. In Section 3.3, we introduce the concept of spatial aggregation as a means of constructing patch-based dictionaries. Experimental results demonstrating the effectiveness of the patch basis functions in suppressing the grating lobes are presented in Section 3.4. Section 3.5 describes an estimation theoretic approach for tuning the parameters of the dictionary of patches. Sections 3.6 and 3.7 introduce a general framework for exploiting sparsity in the scene representation along with reconstruction techniques using patch-based models.

3.2 Imaging Geometry and Grating Lobes

Consider the 1-dimensional array geometry depicted in Figure 3.1 (generalization to two dimensions is straightforward). Let L_1, L_2, and D denote the size of the aperture, the size of the scene, and the distance between the aperture and the scene, respectively. We consider the nominal values $L_1 = L_2 = 15$ cm and $D = 30$ cm in our simulations. The inter-element spacing of a uniform linear array comprised of N elements is given by $d = L_1/(N-1)$. Let Ψ be a set that contains the locations of all of the point scatterers in the scene, and $\Gamma(x') : \Psi \to \mathbb{C}$ denote the complex reflection coefficient corresponding to the point scatterer at location x'. The scene response is an N-dimensional vector denoted by $\boldsymbol{r} \in \mathbb{C}^N$, with the n^{th} entry corresponding to the monostatic transceiver located at

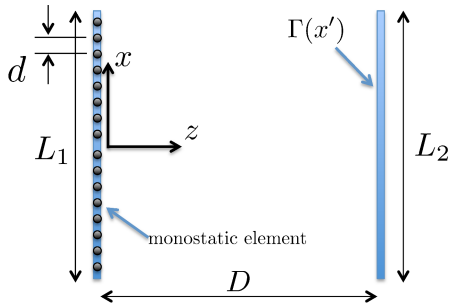


Figure 3.1: Geometry of 1-dimensional monostatic imaging.

x_n, and is given by

$$r[n] = \int_{x' \in \Psi} \Gamma(x') e^{-j2kR(x',x_n)} dx', \tag{3.1}$$

where $k = 2\pi/\lambda$ is the wavenumber, and $R(x', x_n) = \sqrt{D^2 + (x' - x_n)^2}$ is the path length from the transceiver location to the point scatterer at location x'. We use Fresnel approximation [90] (i.e., a first order Taylor approximation) to compute the path lengths, $R(x', x_n) \approx D + \frac{(x'-x_n)^2}{2D}$. This approximation is accurate when D is much larger than L_1 and L_2. The Fresnel diffraction integral, therefore, is given by

$$r[n] \approx e^{-j2kD} \int_{x' \in \Psi} \Gamma(x') e^{-j\frac{k}{D}(x'-x_n)^2} dx'. \tag{3.2}$$

The internal kernel in (3.2) has been studied by Slepian *et al* in the context of time-limited and band-limited functions [61]. It has been shown that the eigenfunctions of this integral equation are prolate spheroidal wave functions (PSWFs), and the corresponding eigenvalues have the interesting property that they remain approximately equal until a critical transition point ($\approx 2\frac{L_1 L_2}{D\lambda}$) where they rapidly decay to zero. The number of non-zero eigenvalues determines the number of degrees of freedom of the imaging system (constrained by the geometry and the wavelength), and is well approximated by

$$DoF \approx 2\frac{L_1 L_2}{D\lambda}. \tag{3.3}$$

In other words, DoF identifies the maximum number of linearly independent measurements of $\Gamma(x')$ that can be achieved using a monostatic array of elements under Fresnel approximation. Increasing the number of array elements beyond DoF can improve the SNR, but it does not lead to new information about the scene. In the next subsection, we take a more practical viewpoint and discuss the significance of the DoF argument in designing sparse monostatic arrays while avoiding grating lobes. It is worth mentioning that the seminal work of Slepian on PSWFs has been applied in various engineering problems, in particular in the context of diffraction limited optics [91], and line-of-sight MIMO communications [27].

3.2.1 Correlation analysis for point-scatterers

Grating lobes appear when two spatially separated point scatterers in the scene generate highly correlated responses. Let us consider two point scatterers located at x'_1 and x'_2 (i.e., $\Psi_i = \{x'_i\}$ for $i \in \{1, 2\}$), with responses denoted by $\boldsymbol{r}_1$ and $\boldsymbol{r}_2$, respectively. The correlation between the two responses is given by

$$\begin{aligned}
\boldsymbol{r}_1^H \boldsymbol{r}_2 \;&\approx \sum_{n=1}^{N} \Gamma_1^* \Gamma_2 \; e^{j\frac{k}{D}(x'_1 - x_n)^2} e^{-j\frac{k}{D}(x'_2 - x_n)^2} \\
&= \Gamma_1^* \Gamma_2 \; e^{j\frac{k}{D}(x'^2_1 - x'^2_2)} \sum_{n=1}^{N} e^{j\frac{2k}{D}(x'_2 - x'_1)x_n} \\
&= c_1 \; \frac{\sin(N\rho/2)}{\sin(\rho/2)} = c_1 \; \mathrm{Dir}(\rho),
\end{aligned} \tag{3.4}$$

where c_1 is a complex coefficient with $|c_1| = |\Gamma_1||\Gamma_2|$, and $\mathrm{Dir}(\rho) = \frac{\sin(N\rho/2)}{\sin(\rho/2)}$ is the well-known Dirichlet kernel with $\rho \triangleq \frac{2k}{D}(x'_2 - x'_1)d$. Note that $\rho \in [0, 2kdL_2/D]$; that is, it takes values in an interval that depends on the signal wavelength and the geometry of the imaging problem. Additionally, $||\boldsymbol{r}_i|| = |\Gamma_i|\sqrt{N}$ for $i \in \{1, 2\}$, so that the normalized

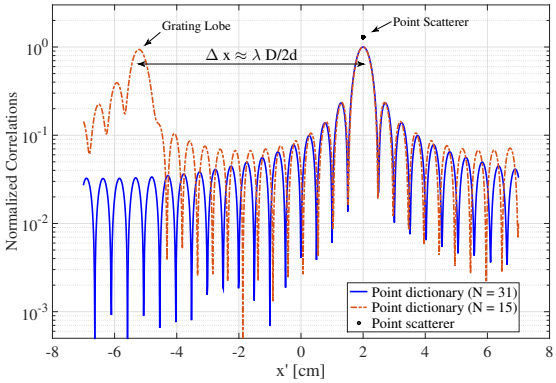


Figure 3.2: Grating lobe artifact for sparse monostatic array.

correlation function is given by

$$\frac{|\boldsymbol{r}_1^H \boldsymbol{r}_2|}{||\boldsymbol{r}_1||.||\boldsymbol{r}_2||} = \frac{1}{N}\text{Dir}(\rho). \tag{3.5}$$

The Dirichlet kernel is a periodic function with period 2π, hence the condition for avoiding aliasing, or grating lobes, translates to constraining the visible range of ρ to be less than 2π, or $\frac{2kdL_2}{D} \leq 2\pi$. This yields $N \geq \frac{2L_1L_2}{D\lambda} + 1$ or $N \geq DoF + 1$. For our nominal parameter values, we obtain that $N \geq 31$.

Figure 3.2 shows the magnitude of the normalized correlations (also known as the ambiguity function) for point scatterers $x_1' = 2$cm fixed, and x_2' varying in the interval $[-7.5, 7.5]$ cm, for $N = 15$ and $N = 31$. We see that setting $N = 15$ leads to a grating lobe artifact at $x_{\text{gl}} \approx -5.5$ cm. The separation between the true point location and the grating lobe is calculated by setting $\rho = 2m\pi$, for $m \in \mathbb{Z}$. It is easy to see that the first grating lobe is at distance $\Delta x = |x_{\text{gl}} - x_1'| \approx \frac{\lambda D}{2d}$ from the true point location (e.g. $\Delta x \approx 7.5$ cm for $N = 15$).

Another important observation is that increasing the number of array elements beyond $\frac{2L_1L_2}{\lambda D} + 1$ has little impact on the normalized correlation function. This is because,

for $N \gg DoF$, we have

$$\begin{aligned}
\frac{|r_1^H r_2|}{||r_1||.||r_2||} &= \frac{1}{N}\frac{\sin\left(\frac{kL_1N}{D(N-1)}(x_2'-x_1')\right)}{\sin\left(\frac{kL_1}{D(N-1)}(x_2'-x_1')\right)} \\
&\approx \frac{1}{N}\frac{\sin\left(\frac{kL_1}{D}(x_2'-x_1')\right)}{\frac{kL_1}{D(N-1)}(x_2'-x_1')} \\
&\approx \operatorname{sinc}\left(\frac{kL_1}{\pi D}(x_2'-x_1')\right) \qquad (3.6)
\end{aligned}$$

where $\operatorname{sinc}(x) \triangleq \frac{\sin(\pi x)}{\pi x}$. Note that increasing the number of array elements beyond DoF does not improve the ambiguity function for locating a point scatterer in the scene, and only leads to an increase in the effective SNR.

3.3 Spatial Aggregation and the Patch Model

We now introduce *spatial aggregation* as a technique for constructing new dictionaries for reconstruction [28, 92]. The idea is to replace the point scatterer as the basic primitive for explaining the scene, with a (continuous) collection of point scatterers, adjacent to one another and having roughly constant reflection coefficient. Such a "patch" primitive is a natural model for explaining extended objects with bounded variation in the reflectivity function $\Gamma(x')$ across space; that is, for scenes where the reflectivity is a *spatially lowpass* function.

Let us consider two collections of point scatterers Ψ_1 and Ψ_2, with their corresponding responses denoted by $\tilde{r}_1$ and $\tilde{r}_2$, respectively. The correlation between the two responses

is calculated as follows:

$$\begin{aligned}
&\tilde{r}_1^H \tilde{r}_2 = \\
&\sum_{n=1}^{N} \int_{x_1' \in \Psi_1} \Gamma^*(x_1') e^{j2kR(x_1', x_n)} dx_1' \int_{x_2' \in \Psi_2} \Gamma(x_2') e^{-j2kR(x_2', x_n)} dx_2' \\
&\approx \iint_{\substack{x_1' \in \Psi_1 \\ x_2' \in \Psi_2}} \Gamma^*(x_1') \Gamma(x_2') e^{j\frac{k}{D}(x_1'^2 - x_2'^2)} \sum_{n=1}^{N} e^{j\frac{2k}{D}(x_2' - x_1')x_n} dx_2' dx_1' \\
&= c_2 \iint_{\substack{x_1' \in \Psi_1 \\ x_2' \in \Psi_2}} \Gamma^*(x_1') \Gamma(x_2') H(x_1', x_2') \mathrm{Dir}(\rho) dx_2' dx_1', \qquad (3.7)
\end{aligned}$$

where $H(x_1', x_2') \triangleq e^{j\frac{k}{D}(x_1'^2 - x_2'^2)}$ is the Spatial Aggregation (SA) kernel. The magnitude of the SA kernel is constant, and the phase is a non-linear function of x_1' and x_2'. It is easy to see that, since the SA kernel is not a function of N, the normalized correlation for the response of Ψ_1 and Ψ_2 is also independent of N for $N \gg DoF$. Figure 3.3 shows the magnitude of the Dirichlet kernel as well as the phase of the product $H(x_1', x_2')\mathrm{Dir}(\frac{2k}{D}(x_2' - x_1')d)$ for $x_1', x_2' \in [-7.5, 7.5]$ cm. The key observation is that the phase of SA kernel is nearly constant across the main lobe, while exhibiting rapid variations across the grating lobes. This incoherency induced by the SA kernel is what leads to significant suppression of grating lobes. In order to illustrate the effect of spatial aggregation, consider a simple dictionary of fixed size intervals, defined by $\mathcal{P}_w \triangleq \{\Psi = [\alpha - w/2, \alpha + w/2] : \alpha \in [-L_2/2, L_2/2]\}$. We term this a dictionary of *patches* of w cm width. For the special case of $w \to 0$, we obtain the dictionary of point scatterers. Figure 3.4 shows the magnitude of normalized correlations for $\Psi_1 = [1, 3]$ cm fixed (2cm width), and $\Psi_2 \in \mathcal{P}_1$ for $N = 15$. We also plot the correlations of Ψ_1 response and the dictionary of point scatterers $\mathcal{P}_0$ for comparison. We see that the effect of SA is three-fold: (1) suppressing the grating lobe, (2) reducing the side lobe level, and (3) widening the main lobe. It is important to note

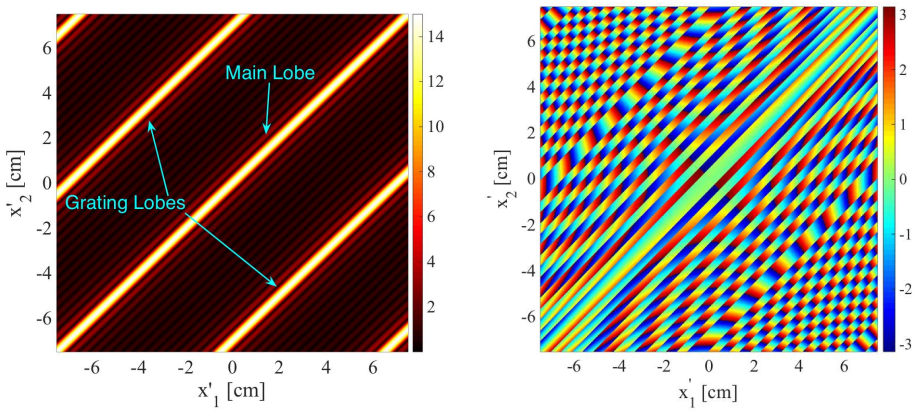


Figure 3.3: (left) Magnitude of Dirichlet ($N = 15$), (right) Phase of the product of SA kernel and Dirichlet kernel $H(x_1', x_2')\text{Dir}(\rho)$.

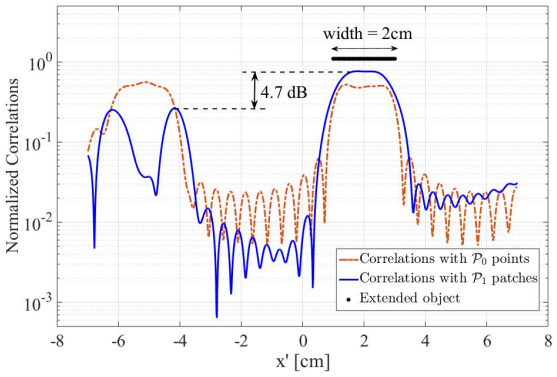


Figure 3.4: Reduction of grating/side lobes by spatial aggregation.

that the main lobe of patch correlations is representing an extended object, therefore, its width should be analyzed with respect to the width of the patch itself.

3.4 Experimental Results

Our hardware testbed is a 60 GHz continuous-wave quasi-monostatic (transmit and receive antennas are slightly separated, but approximately appear to be co-located as viewed from the target) radar transceiver, equipped with dual high-gain horn antennas [93]. We use a mechanical platform to move the imager on a plane parallel to the scene,

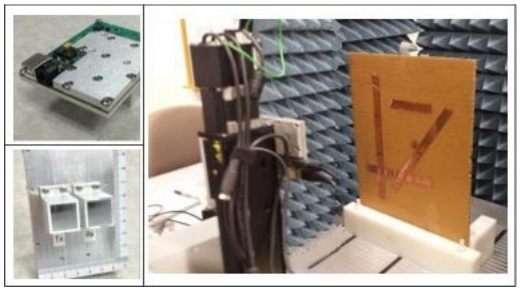


Figure 3.5: Experimental data collection using 60 GHz quasi-monostatic radar system.

thereby emulating a 2D array of sensors. The travel distance of the imager is of the order of the form factor of a portable handheld device (covering a 15 cm $\times$ 15 cm area).

3.4.1 2D reconstruction in the spatial domain

In this subsection, we consider two uniform planar array configurations: (1) dense array of 30 $\times$ 30 elements (i.e. $d \approx \lambda = 0.5$ cm), (2) sparse array of 15 $\times$ 15 elements (i.e. $d \approx 2\lambda = 1$ cm). At each step of movement, the scene response is measured using a single frequency continuous-wave waveform at 60 GHz, and stored in a vector. We consider a scene of copper strips that is placed parallel to the array at a distance $D = 30$ cm (Figure 3.5).

The first reconstruction method that we consider is standard SAR processing, which may be viewed as matched filtering (MF) with respect to a point scatterer based signal model [94]. The second approach is based on computing the correlation of the measured response with the responses of 1.5 cm $\times$ 1.5 cm square patches; that is, the collection of patches obtained by sliding a 1.5 cm $\times$ 1.5 cm window over the entire scene. We refer to this approach as patch-based MF. Figure 3.6 shows the results of point-based MF. We see that grating lobes lead to significant deterioration in image quality for the sparse array. Figure 3.7 shows the result of patch-based MF. We see significant improvement in the image quality, in terms of suppressing the grating lobes and increasing the dynamic

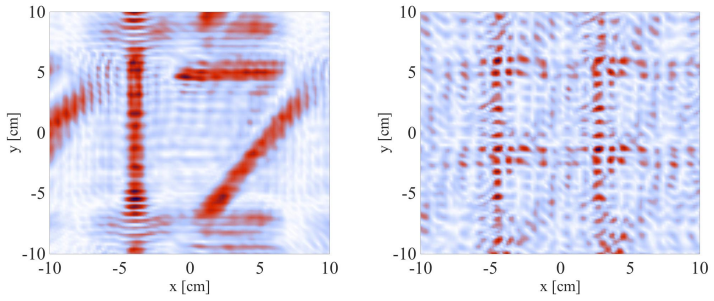


Figure 3.6: Point-based MF (left) dense array (right) sparse array.

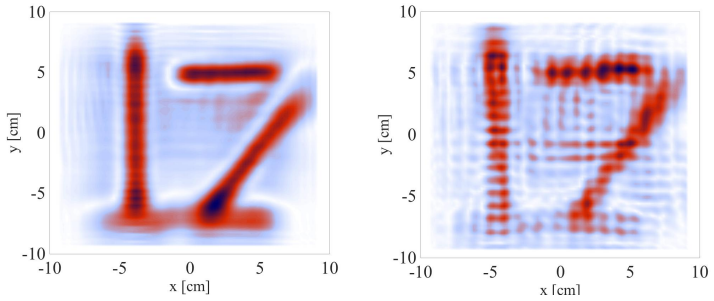


Figure 3.7: Patch-based MF (left) dense array (right) sparse array.

range, e.g. the horizontal strip at the bottom of the scene becomes visible for the dense array deployment.

As we discussed in Subsection 3.2.1, the first grating lobe appears at distance $\Delta x \approx \frac{\lambda D}{2d}$ from the true point location. This corresponds to $\Delta x \approx 15$ cm and 7.5 cm, for our nominal geometry with $N = 30$ and $N = 15$ element arrays, respectively. The dependency of Δx on the wavelength λ, suggests that by incorporating a wideband signaling scheme, one can potentially suppress the grating lobe effects, due to the incoherency in the location of the grating lobes across different frequencies [81]. Next, we investigate the potential of wideband signaling in suppressing the grating lobes using experimental data for 3D image reconstruction.

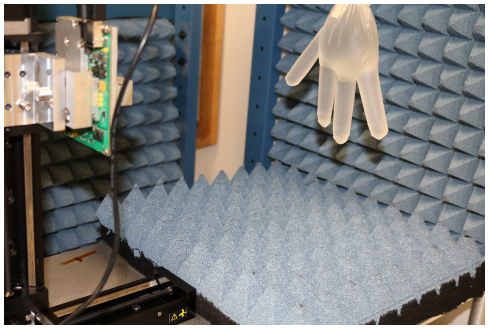


Figure 3.8: Experimental data collection using SFCW radar system.

3.4.2 Wideband SFCW signaling and 3D reconstruction

In this subsection, we investigate wideband Stepped-Frequency Continuous-Wave (SFCW) signaling using a 2D array of quasi-monostatic elements, followed by 3D image reconstruction techniques in the spatial domain. We consider two uniform planar array configurations: (1) dense array of 50×50 elements (i.e. $d \approx 0.6\lambda = 0.3$ cm), (2) sparse array of 17×17 elements (i.e. $d \approx 1.8\lambda = 0.9$ cm). At each step of movement, the scene response is measured in discrete frequency steps, covering 55 to 60 GHz band in $N_\omega = 100$ uniformly spaced intervals. The overall response of scene is saved in a 3D matrix $f(x, y; \omega) \in \mathbb{C}^{N \times N \times N_\omega}$. We consider a scene of a plastic glove (with four fingers) filled with salt water, that is placed parallel to the array at a distance $D \approx 28$ cm, as shown in Figure 3.8.

We apply Matched Filter reconstruction technique across both space (aperture) and frequency (bandwidth) for any hypothesized location in the scene [95]. Figure 3.9 shows the output of MF algorithm for the dictionary of point-scatterers, for both dense and sparse array configurations. The scene information in preserved for the dense array and we are able to identify three fingers and the palm of the hand in the image. The fourth finger, however, is not visible in the reconstructed image. This is due to the beam pattern of the horn antennas (low power illumination of the edges of the scene), as well

as the small radar-cross-section of the finger, which lead to significant power loss for the backscattered electromagnetic wave. For the sparse array configuration, as shown in Figure 3.9(b), point-based MF does not preserve the scene information due to the grating lobes. This indicates that the frequency diversity induced by wideband SFCW signaling is not sufficient for eliminating the grating lobes in sparse monostatic arrays. Next, we apply patch-based MF, which entails computing the correlation of the overall received response across space and frequency, with the dictionary of 1.5 cm $\times$ 1.5 cm square patches at any hypothesized location in the scene. As shown in Figure 3.10, the algorithm reduces grating lobes and preserves information in both dense and sparse array configurations.

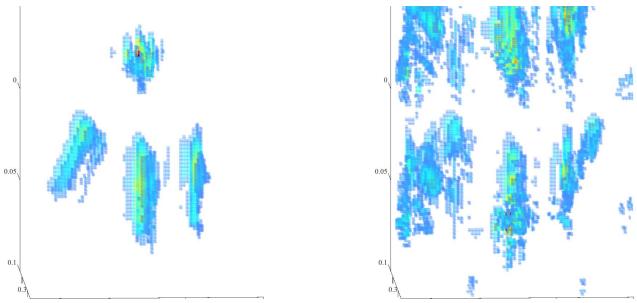


Figure 3.9: Point-based MF reconstruction of hand sample (left) dense array (right) sparse array.

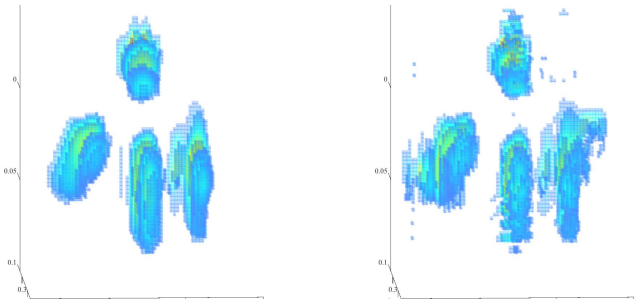


Figure 3.10: Patch-based MF reconstruction of hand sample (left) dense array (right) sparse array.

3.5 Estimation-Theoretic Dictionary Tuning

We wish to construct a dictionary that is well-matched to simple distributed targets, while providing robustness against array sparsity. Ideally, we would like to identify the smallest patch size (in order to get the best possible resolution), that provides adequate grating lobe suppression. To this end, we turn to the Ziv-Zakai bound (ZZB) for estimating the location of a single atom in the presence of Additive White Gaussian Noise (AWGN). The ZZB is a "global" (Bayesian) bound, i.e., it assumes a prior distribution for the unknown parameter, and bounds the mean squared error (MSE) of *any* estimator [63]. It is well-known that ZZB serves as an accurate predictor of best possible estimation performance over a wide range of SNR values. Roughly speaking, ZZB accounts for both coarse-grained errors in which the parameter estimate falls into the "wrong bin" (in our case, this is determined by grating lobes and side lobes), as well as fine-grained errors within the right bin (this is determined by the main lobe).

As the SNR increases, a point is reached where (on a log-log scale), the ZZB decreases linearly with SNR. After this point, the maximum likelihood estimate of the parameter, with high probability, is close to the true parameter value [63] (i.e., it is in the "right bin"). This threshold SNR, termed the ZZB threshold, indicates a region of operation known as the "asymptotic" regime, where the estimation performance is only limited by main lobe ambiguity (i.e., neither noise nor grating lobes can cause large estimation errors). In this regime, the ZZB is close to the Cramér-Rao Bound (CRB), which operates under the assumption that we are in the right bin. For a sparse array, our goal is to tune our dictionary parameters (i.e., the patch size) so that the ZZB threshold does not exhibit significant deterioration relative to a system with as many array elements as the DoF.

Figure 3.11 plots the ZZB for estimating the location of a point scatterer for different values of N, the number of array elements. We see that the bounds are indistinguish-

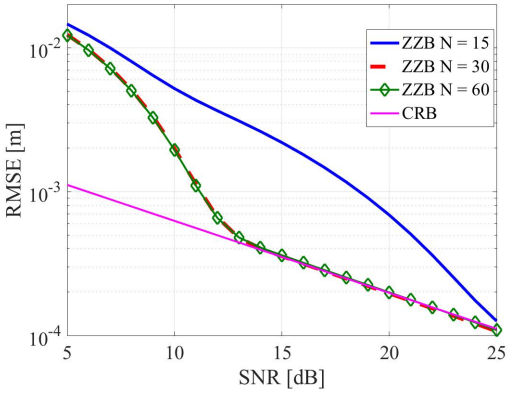


Figure 3.11: ZZB for the dictionary of point-scatterers.

able for $N = 30$ and $N = 60$. This is a consequence of our previous observation that the normalized correlation function does not change much when we increase N beyond $DoF \approx 30$ elements (see Section 3.2). On the other hand, decreasing the number of elements to $N = 15$ introduces grating lobes. The existence of the grating lobe is reflected in the corresponding ZZB curve where the threshold SNR has increased by about 13 dB compared to the dense arrays ($N = 30$ and 60). This significant increase in the ZZB threshold is because the multi-modal structure of the correlation function leads to a fundamental ambiguity in estimating the location of a single point scatterer.

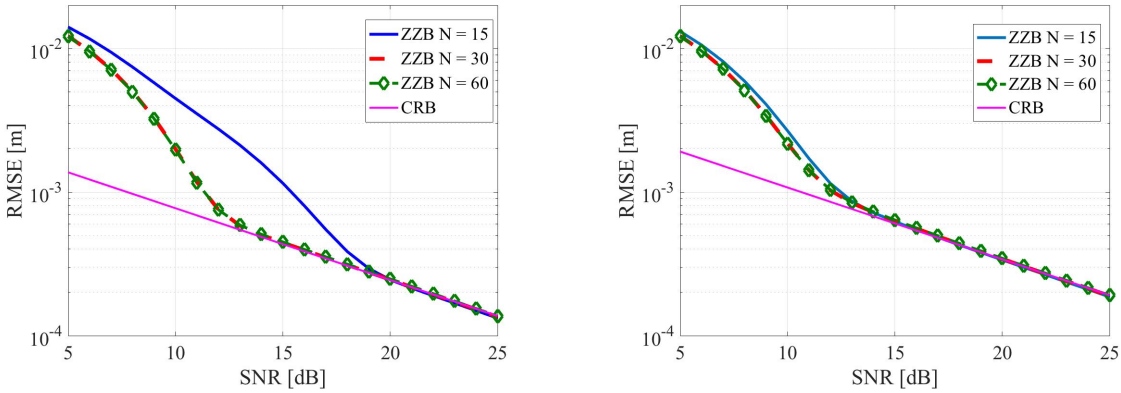


Figure 3.12: ZZB for estimating the location of a single patch in the dictionary of (left) patches of size $w = 0.5$ cm (right) patches of size $w = 1$ cm.

Now, consider the patch-based dictionary $\mathcal{P}_w$ for patch width $w > 0$. Our goal is to tune the parameter w such that its corresponding dictionary of patches suppresses grating lobes sufficiently for a sparse array with a given number of elements. To this end, for each candidate w, we compute the ZZB for estimating the location of a single patch given that the size of the patch is known *a priori*. The gap between the ZZB thresholds corresponding to the sparse array and a dense array is used as an indicator to verify whether the current dictionary is robust against the sparsity of the array, or whether it leads to a multi-modal correlation structure. Figure 3.12 shows that setting $w = 0.5$ cm leads to a 6 dB gap between the SNR thresholds corresponding to $N = 15$ and $N = 30$, whereas setting $w = 1$ cm eliminates this gap. Thus, for this particular scene geometry, array form factor, and number of elements ($N = 15$), a patch width of 1 cm is a better choice. This approach can be generalized to 2D arrays and more complicated parameterized dictionaries of spatially extended atoms. In the next section, we investigate a new technique for image reconstruction that leverages the "sparse representation" of the scene in the dictionary of patches. This parsimonious signal model, with appropriately designed estimation algorithms, allows us to "super-resolve" beyond the limits of conventional radar theory [83].

3.6 Sparsity-Driven Imaging

Simple scenes usually admit a sparse representation in a dictionary of spatially-extended objects. For example, each finger in the image of the plastic glove in Section 3.4.2 can be approximated by a few concatenated patches. Such sparse representations have multiple potential advantages: (1) as pointed out in [84], sparse techniques for image formation can increase resolvability of targets, facilitate segmentation, and provide robustness to limitations in data quality and quantity; (2) sparse representations pro-

vide a framework for analyzing the scene at an "information rate" that is potentially significantly lower than the Nyquist rate [96], allowing minimalistic encoding of scene information; (3) they enable super-resolution via estimation of a small number of parameters associated with the representation (patches in our case); and (4) they may provide a basis for developing efficient algorithms for detecting scene changes via tracking the parameters of the estimated patches.

We formulate image formation as sparse reconstruction [13], assuming that we can approximate the scene response by a linear combination of a few atoms from a predefined dictionary. The response of each atom in the dictionary is represented by a κ-dimensional vector $\boldsymbol{r} \in \mathbb{C}^{\kappa}$. For a dictionary of ζ atoms, we construct a matrix $R \in \mathbb{C}^{\kappa \times \zeta}$, whose columns are the responses of its atoms. That is, $R = [\boldsymbol{r}_1 \ \boldsymbol{r}_2 \ \ldots \ \boldsymbol{r}_\zeta]$. The scene response is represented by

$$\mathbf{y} = R\mathbf{g} + \mathbf{z}, \tag{3.8}$$

where $\mathbf{g} \in \mathbb{C}^{\kappa \times 1}$ includes the complex gains corresponding to each atom, and $\mathbf{z} \sim \mathcal{CN}(0, \sigma^2 \mathbb{I}_\kappa)$ is the AWGN. Let $||\cdot||_0$ denote the counting function (also known as ℓ_0 norm), which returns the number of nonzero elements of its input vector. Sparsity-driven imaging refers to the setting in which $||\mathbf{g}||_0 \ll \zeta$; that is, the scene admits a *sparse representation* in the constructed dictionary.

Our goal in sparse reconstruction framework is to find the *maximally sparse* representation of the scene, while allowing for some error tolerance $\epsilon \geq 0$ due to noise and modeling errors, by solving the following combinatorial optimization problem,

$$\underset{\mathbf{g}}{\text{minimize}} \ ||\mathbf{g}||_0 \quad \text{subject to} \quad ||R\mathbf{g} - \mathbf{y}||_2 < \epsilon. \tag{3.9}$$

Finding the exact solution of (3.9) without any constraints on the matrix R is known

to be NP-hard [13]. Therefore, we resort to computationally tractable algorithms that generate approximate solutions. Specifically, we focus on *Convex Relaxation* and *Greedy Pursuit*, two of the most popular techniques used in practice, which also offer certain theoretical guarantees of their performance [97, 14].

3.6.1 Convex relaxation

A popular approach for sparse reconstruction is to replace the ℓ_0 norm in the optimization problem (3.9), with ℓ_1 norm, resulting in a convex optimization program,

$$\underset{\mathbf{g}}{\text{minimize}} \; ||\mathbf{g}||_1 \quad \text{subject to} \quad ||R\mathbf{g} - \mathbf{y}||_2 < \epsilon. \tag{3.10}$$

Note that the ℓ_1 norm is the closest convex function to the ℓ_0 function. It has been shown that under suitable conditions on the matrix R, and when the optimal $\mathbf{g}$ is reasonably sparse, then this convex relaxation leads to the exact solution of the original problem in (3.9) [97]. One can also incorporate the ℓ_2-error constraint in (3.10) as part of the objective function, yielding a scalarized dual-objective optimization program,

$$\underset{\mathbf{g}}{\text{minimize}} \; \frac{1}{2}||R\mathbf{g} - \mathbf{y}||_2^2 + \lambda||\mathbf{g}||_1, \tag{3.11}$$

where the regularization parameter $\lambda > 0$ balances the two objectives of minimizing the residual squared error (ℓ_2 term) and sparsity (ℓ_1 term). Increasing the value of λ typically leads to sparser solutions. Let $\lambda_{max} \triangleq ||R^H\mathbf{y}||_\infty$. Setting $\lambda > \lambda_{max}$ leads to $\mathbf{g} = \mathbf{0}$ as the solution of (3.11). The formulation in (3.11) is also known as least absolute shrinkage and selection operator (LASSO), which was first introduced in the context of feature selection [89]. We apply LASSO to reconstruct the scene of copper strips (Figure 3.5), given the sparse 15×15 element array configuration (described in Section 3.4.1). Figure 3.13 shows

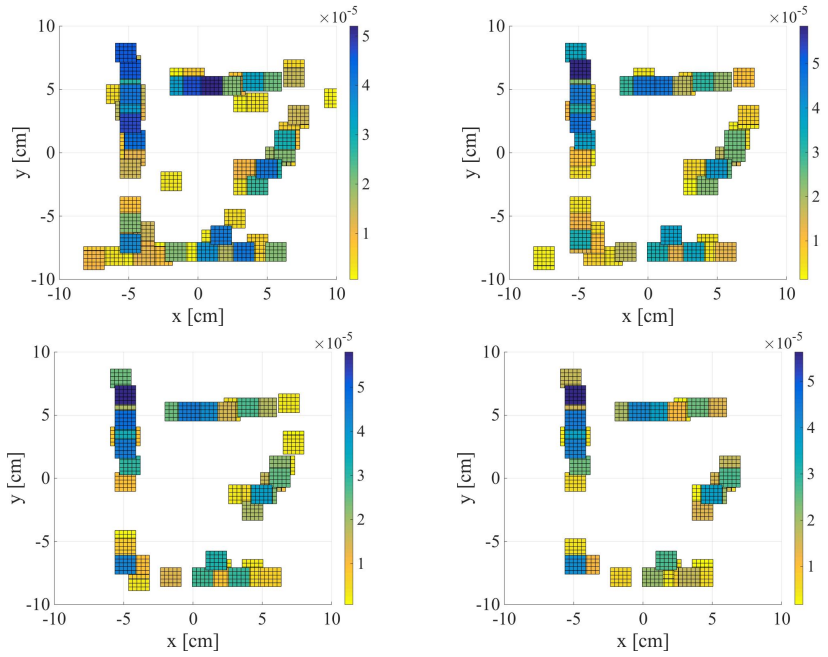


Figure 3.13: Sparse reconstruction of the scene of copper strips (shown in Figure 3.5), using LASSO for different values of λ: (up left) $\lambda = 0.25 \times \lambda_{max}$, (up right) $\lambda = 0.3 \times \lambda_{max}$, (down left) $\lambda = 0.35 \times \lambda_{max}$, (down right) $\lambda = 0.4 \times \lambda_{max}$.

LASSO outputs for different values of λ, when the matrix R is constructed based on the dictionary of 1.5 cm $\times$ 1.5 cm square patches. The color of each patch is proportional to its gain. We see that increasing λ leads to sparser solutions, while preserving most of the information in the scene. However, due to the modeling errors in approximating the scene by square patches of a fixed size (with constant reflectivity across each patch), we see significant overlap of multiple patches trying to explain the reflectivity of the actual scene. As we show next, greedy approaches to sparse reconstruction work better.

3.6.2 Greedy pursuit

Another approach for finding the maximally sparse representation of the scene in the dictionary of patches is based on Greedy Pursuit [13]. NOMP algorithm falls in this category and is applicable for sparse approximation in any continuously parametrized

overcomplete dictionary. Greedy iterative approaches (e.g., OMP and NOMP) are particularly attractive due to their low computational complexity and ease of implementation [13].

For a scene containing $\mathcal{K}$ patches, the overall response is modeled as

$$\mathbf{y} = \sum_{i=1}^{\mathcal{K}} g_{\alpha_i} \boldsymbol{r}_{\alpha_i} + \mathbf{z}, \tag{3.12}$$

where $g_{\alpha_i} \in \mathbb{C}$ denotes the complex gain for i^{th} patch. We now present a high-level description of NOMP for sparse recovery in our radar imaging problem. Let Λ be the set of detected patches. In each iteration of the algorithm, an atom that yields the greatest improvement in the approximation quality is identified and added to Λ. After that, a *cyclic refinement* step is applied to all the atoms in Λ (i.e., to all of the atoms that have been estimated in the previous iteration), therefore giving them a chance to re-evaluate their estimates to incorporate the effect of newly detected atom. We do not make *a priori* assumptions on the number of patches, and set the stopping criterion based on the relative energy reduction of the residual signal (i.e., the portion of the signal not explained by the currently estimated set of patches). We stop looking for further patches when the relative energy reduction of the residual goes below a threshold, denoted by ϵ. The iterative sparse reconstruction algorithm is stated as follows:

1. Let $\boldsymbol{q}_0 = \mathbf{y}$, $\varepsilon_{q_0} = ||\boldsymbol{q}_0||^2$, and loop counter $i \leftarrow 1$.

2. Find $\lambda = \arg\max_\beta \{\frac{|\boldsymbol{q}_{i-1}^H \boldsymbol{r}_\beta|}{||\boldsymbol{q}_{i-1}||.||\boldsymbol{r}_\beta||} : \boldsymbol{r}_\beta \in R\}$. Set $g_\lambda = \frac{(\boldsymbol{r}_\lambda^H \boldsymbol{q}_{i-1})}{||\boldsymbol{r}_\lambda||^2}$, and update $\Lambda \leftarrow \Lambda \cup \{(\lambda, g_\lambda)\}$.

3. Cyclicly refine centers/gains for all patches in Λ.

4. Update all gains by least squares for best approximation of $\mathbf{y}$ with the atoms chosen

so far.

5. Find the new residual $\boldsymbol{q}_i = \mathbf{y} - \sum_{\lambda \in \Lambda} g_\lambda \boldsymbol{r}_\lambda$, and compute its energy $\varepsilon_{q_i} = ||\boldsymbol{q}_i||^2$.

6. If $|\varepsilon_{q_i} - \varepsilon_{q_{i-1}}|/\varepsilon_{q_0} > \epsilon$, then set $i \leftarrow i+1$, and go back to Step 2, otherwise, declare Λ as the output of the program.

Figure 3.14 shows the result of the patch detection algorithm for the same scene of copper strips (Figure 3.5), with only 15×15 equispaced emulated array elements (sparse array). The dictionary R contains the 1.5 cm $\times$ 1.5 cm square patches. We see that the algorithm is able to detect the horizontal strip at the bottom of the scene, despite the sparsity of the array, which is a significant improvement over the MF processing results in Figures 3.6(b) and 3.7(b). Moreover, in comparison to LASSO output in Figure 3.13, we see that NOMP is more immune to the modeling errors, and is able to generate a sparse approximation of the scene with minimal overlap of the detected patches.

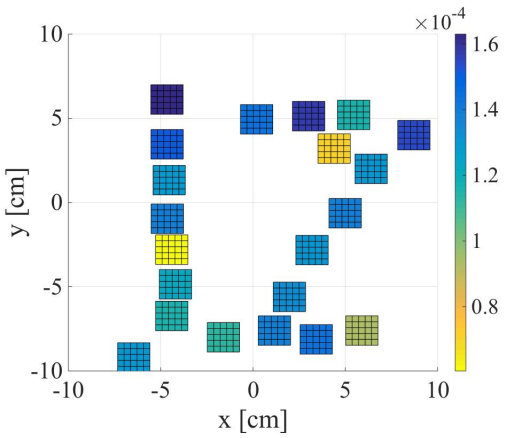


Figure 3.14: Image reconstruction using NOMP algorithm in the sparse array configuration. The patch sizes are fixed (1.5cm×1.5cm), and we only refine the location of patches.

3.7 Generalizations

We now briefly discuss potential avenues for generalizing the proposed framework. In the construction of the dictionary of spatially extended atoms, we explicitly incorporate information about the array geometry in computing the response for each atom. This approach allows for the design of a dictionary that is well-matched to both the geometry of the array, and the nature of the scene being imaged, and can be easily extended to other settings, such as non-uniform arrays.

3.7.1 Dynamic dictionaries

One of the problems associated with scene-based dictionary design is that the number of atoms in the dictionary may become excessively large in order to capture various shapes and sizes of the targets in the scene. This may lead to computational bottlenecks, especially for real-time applications. One approach to alleviate this problem is to construct a dynamic dictionary by bootstrapping from a base dictionary characterized with a few parameters, with a relatively small number of atoms. Additional atoms could then be added to this base dictionary during image reconstruction. For example, the dictionary of fixed-size square patches employed in Section 3.6 can be augmented by allowing modification of patch sizes to better approximate the response to a given scene. This is easily accomplished by refining the width of patches along with their centers and gains in Step 3 of the NOMP algorithm. Figure 3.15 shows the output of NOMP where we start with the dictionary of 1.5 cm$\times$1.5 cm patches as the base, and then refine the sizes of the detected patches throughout the reconstruction process.

As another example, consider a dictionary of circular patches. Figure 3.16 shows the output of NOMP algorithm where we use the collection of 1cm radius atoms as the base dictionary, but then allow for both center and radius refinements. We see that the

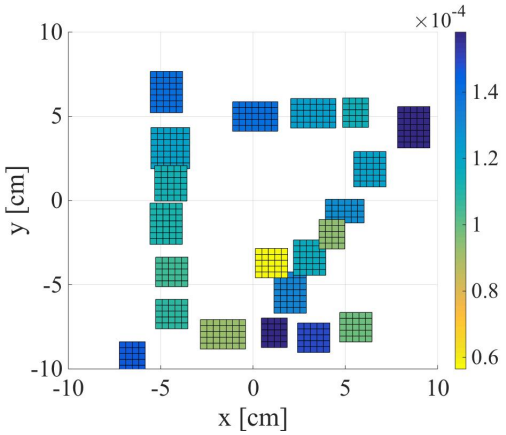


Figure 3.15: Image reconstruction using NOMP algorithm in the sparse array configuration. We refine both the location and size of rectangular patches.

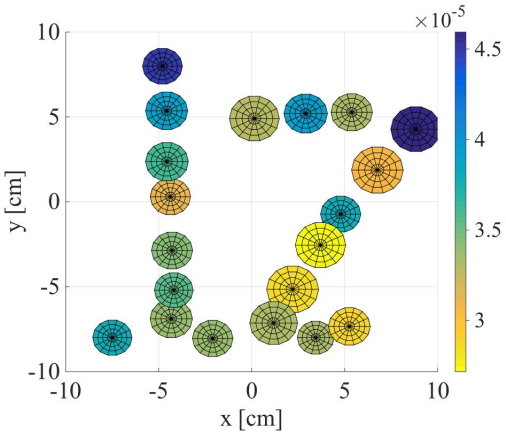


Figure 3.16: Image reconstruction using NOMP algorithm in the sparse array configuration. We refine both the location and size of circular patches.

general structure of the scene is reflected in the image, including the horizontal strip at the bottom of the scene. An advantage of the circular dictionary is that the spatial size of the atoms is controlled by a single parameter (radius), hence dynamic adaptation is more computationally efficient compared to reconstruction using a dictionary of square patches.

Chapter 4

Degrees of Freedom Analysis of Imaging Systems

4.1 Introduction

In this chapter, we take a more fundamental and systematic approach based on singular decomposition of the imaging system, to understand the information capacity and the limits of performance for various geometries. In particular, we want to identify the number of degrees of freedom of an imaging system, i.e., the number of independent parameter that can be inferred from an unknown scene using an array of antenna elements. It is also of great interest to determine whether there are significant DoF gains to be obtained using multistatic architecture compare to monostatic. The significance of the DoF analysis has multiple theoretical and practical aspects: (i) it is related to fundamental performance measures such as achievable resolution and the information capacity of the system, (ii) it provides guidelines to design efficient and practical array architectures under various cost and complexity constraints, (iii) and it also provides crucial insights on the performance of different image reconstruction algorithms.

4.1.1 Related work

Classical theories for DoF analysis of imaging systems stem from applying Shannon sampling theorem. In a series of fundamental papers, Di Francia derived the significant conclusion that an image formed through a finite size aperture has a finite number of DoF [29, 98]. Since there is no limitation on the number of DoF of the scene, it follows that many different scenes can have exactly the same image. However, this result is not mathematically correct, as has been pointed out by multiple authors [99, 100]. The reason being that if the scene is of finite size, then the knowledge of its Fourier transform over a bounded domain is enough to reconstruct it exactly by using analytic continuation. The solution to this mathematical argument lies in the inevitable existence of noise in practical system, and has been clarified by introducing the notion of *effective* or *practically useful* DoF [24], by applying the seminal work of Slepian, Landau, and Pollak on prolate spheroidal wave functions (PSWFs) [61, 101]. The PSWFs theory shows that the eigenvalues corresponding to a finite Fourier integral operator are approximately constant up until a critical point, after which they decay exponentially to zero. Consequently, in the presence of noise, only a finite number of eigenvalues can be used to accurately determine the output of the integral operator.

The PSWFs theory can be directly applied to the geometry of symmetric parallel planner surfaces at far field, where we can use Fresnel approximation [102] for analyzing the measurement model [103, 104, 105]. Consequently, it has been shown that the number of DoF of such imaging systems is also finite, and the corresponding eigenfunctions are related to the PSWFs. Similar techniques have been applied to analyze multiple observation domains [106], and orthogonal planes geometry [107], nevertheless, they do not directly generalize to short range, asymmetric or tilted planes geometries. The authors in [25] present a general theory for computing the electromagnet DoF of optical systems

under arbitrary boundary conditions, following a sequential optimization framework for finding the strongest connected source and receiving functions that span the input and output spaces, respectively. This approach is equivalent to the SVD of the system integral operator, and it shows that the number of practically useful DoF are essentially finite under general boundary conditions. In this chapter, we use numerical SVD computations to verify the accuracy of our SBP analysis for various geometries.

The terms space-bandwidth product, defined as the area within the Wigner distribution representation, has been used in the literature for evaluating the information content of optical signals and systems [108, 109]. The Wigner distribution forms an intermediate signal description between the pure spatial representation and the pure Fourier domain representation, and can be roughly interpreted as the local spatial spectrum of the signal [110]. The Wigner distribution has been derived for a 1D signal and its corresponding 1D Fourier Transform, and does not directly apply to our measurement model where the desired information is seen through an electromagnetic imaging system. We first evaluate the measured k-space spectrum for any point scatterer in the scene, then compute the corresponding spatial frequency bandwidth *after removing possible redundancies*.

The rest of the chapter is organized as follows: Section 4.2 presents the measurement model and mathematical background for SVD analysis and the k-space representation of the system. Sections 4.3 investigates the k-space spectrum corresponding to a point scatterer seen through monostatic and multistatic array architectures. In Section 4.4, we go through the details of SBP computations for different imaging geometries, and verify the accuracy of the results by numerical SVD computations. Section 4.5 investigates the implications of Fresnel approximation for parallel planes geometry; it shows that SBP computations converge to approximate solutions of previous models, and provides insights for the design of multistatic arrays in the Fresnel zone. In Section 4.6 we investigate image formation techniques that aim to solve the inverse scattering problem, and analyze the

achievable resolution for monostatic and multistatic arrays.

4.2 Mathematical Model and Background

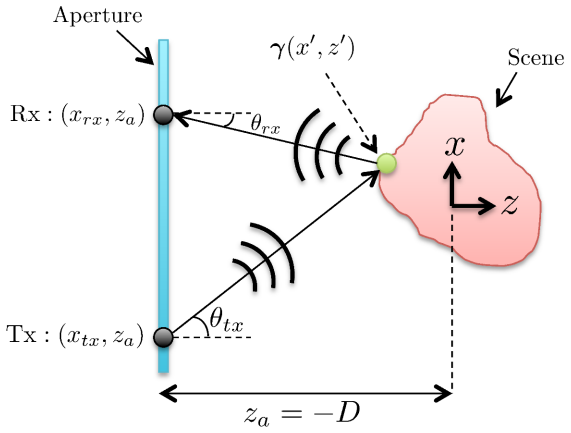


Figure 4.1: Geometry of 1-dimensional bistatic pair at distance $z_a = -D$ from the center of the scene.

Consider the 1-dimensional (1D) aperture depicted in Figure 4.1. The scene (object) is located at the origin, with the corresponding reflectivity function defined by $\boldsymbol{\gamma} : \mathcal{A} \to \mathbb{C}$, where $\mathcal{A} \subset \mathbb{R}^2$ is bounded. We restrict the Tx and Rx antenna elements to be located on the same plane characterized by $z_a = -D$. Assume the Tx element located at (x_{tx}, z_a) illuminates the entire scene, and the Rx element located at (x_{rx}, z_a), measures the back-scattered signal. The observed signal over an aperture of length L_1 is denoted by $\boldsymbol{s} : \mathcal{B} \to \mathbb{C}$, where $\mathcal{B} = [-L_1/2, L_1/2] \times [-L_1/2, L_1/2]$. The relationship between the scene reflectivity function and the observed signal over the aperture is governed by Helmholtz wave equation [111]. Under Born approximation [30], the solution of scalar Helmholtz equation, assuming homogenous isotropic media (and simplified by dropping the space attenuation factors), boils down to the following *linear* integral equation,

$$\boldsymbol{s}(x_{tx}, x_{rx}) = \int_{\mathcal{A}} \xi(x_{tx}, x_{rx}, x', z') \boldsymbol{\gamma}(x', z') dx' dz', \tag{4.1}$$

where,

$$\xi(x_{tx}, x_{rx}, x', z') = e^{-jkR(x_{tx},x';z')} e^{-jkR(x_{rx},x';z')}, \tag{4.2}$$

denotes the space-variant impulse response of the system, and $R(x, x'; z') \triangleq \sqrt{(x - x')^{+}(z_a - z')^2}$ is the Euclidean distance between the point (x, z_a) on the aperture plane, and the point-scatterer in the scene located at (x', z'). The wavenumber is denoted by $k = \frac{2\pi}{\lambda}$, where λ is the signal wavelength. We assume $\boldsymbol{\gamma} \in \Psi$ and $s \in \Phi$, where $\Psi \triangleq \mathcal{L}^2(\mathcal{A})$ and $\Phi \triangleq \mathcal{L}^2(\mathcal{B})$ represent the Hilbert spaces of square integrable functions over $\mathcal{A}$ and $\mathcal{B}$, respectively. This assumption places a (physically plausible) restriction on the scene reflectivity function and the scattered electromagnetic fields to have finite energy values. It is convenient to recast the linear observation model in the following operator form,

$$\boldsymbol{s} = \Xi\boldsymbol{\gamma}, \tag{4.3}$$

where $\Xi : \Psi \to \Phi$ is defined by the right-hand side of (4.1). It is easy to see that the integral kernel (4.2) satisfies

$$\iint_{\mathcal{B}\mathcal{A}} |\xi(x_{tx}, x_{rx}, x', z')|^2 dx_{tx} dx_{rx} dx' dz' < \infty, \tag{4.4}$$

hence, Ξ belongs to the class of Hilbert-Schmidt operators, and is *compact* [112]. By virtue of linearity and compactness, we can invoke the Spectral Theorem [112, 113], and introduce the singular value decomposition of Ξ, denoted by $\{\sigma_i, \psi_i, \phi_i\}_{i\in\mathbb{N}}$, where $\sigma_i \in \mathbb{R}^{+}$are the singular values, and $\psi_i \in \Psi$ and $\phi_i \in \Phi$ are the right and left singular functions, respectively. The operatoration $\boldsymbol{s} = \Xi\boldsymbol{\gamma}$ can thus be expressed as

$$\boldsymbol{s} = \sum_{i=1}^{\infty} \sigma_i \phi_i \langle \boldsymbol{\gamma}, \psi_i \rangle_{\Psi}, \tag{4.5}$$

where $\langle \cdot, \cdot \rangle_{\Psi}$ denotes the inner-product in Ψ. The SVD of Ξ is equivalent to the following series expansion of the integral kernel in (4.2),

$$\xi(x_{tx}, x_{rx}, x', z') = \sum_{i=1}^{\infty} \sigma_i \phi_i(x_{tx}, x_{rx}) \psi_i^*(x', z'). \tag{4.6}$$

The sets of singular functions $\{\psi_i\}_{i\in\mathbb{N}}$ and $\{\phi_i\}_{i\in\mathbb{N}}$ are orthonormal bases for Ψ and Φ, respectively. From (4.5), we obtain the following one-to-one correspondence between the two sets of singular functions,

$$\phi_i = \frac{1}{\sigma_i} \Xi \psi_i, \quad \forall i. \tag{4.7}$$

Square integrability of the kernel in Equation (4.4) leads to

$$\sum_{i=1}^{\infty} \sigma_i^2 < \infty, \tag{4.8}$$

i.e., the sum of squares of the singular values of a Hilbert-Schmidt operator is convergent [112]. Therefore, if the sequence of singular values is ordered in a non-increasing way, we have $\sigma_i^2 \to 0$ for $i \to \infty$. In other words, although in principle the number of non-zero singular values could be infinite, the number of *practically useful* singular values is finite [24, 25]. The normalized sum of singular values $\bar{\Sigma} = \sum_i (\sigma_i / \sigma_{max})$, where $\sigma_{max} = \max\{\sigma_i\}$, can be associated with the number of degrees of freedom for a system with a steplike behavior of the singular values [31, 101], i.e., all useful non-zero singular values are approximately equal, up to a certain threshold where they decay rapidly to zero. In general, this condition is not satisfied for imaging systems and the singular values decay gradually to zero [25]. Our goal in this chapter is to find a criteria to determine the number of useful non-zero singular values of Ξ for different imaging scenarios. In general, the number of independent parameters that we can extract from an unknown signal using a linear operator is precisely determined by the number of non-zero singular values of

the operator. For our simulations we compute the singular system of the operator Ξ by discretizing the kernel provided by equation (4.2), over the parameter spaces $\mathcal{A}$ and $\mathcal{B}$, and compute the SVD in Matlab. In the next subsection, we review k-space (spatial frequency domain) representation of the integral operation in Equation (4.1), which is a crucial step in defining the SBP of imaging system.

4.2.1 k-space representation

Taking the 2D Fourier Transform (FT) of $\boldsymbol{s}(x_{tx}, x_{rx})$ described by (4.1) and (4.2), over the aperture, yields the data representation in the spatial frequency domain,

$$\begin{aligned} & S(k_{x_{tx}}, k_{x_{rx}}) = \mathrm{FT}_{2D}\{\boldsymbol{s}(x_{tx}, x_{rx})\} \\ \triangleq & \iint\limits_{x_{tx}x_{rx}} \boldsymbol{s}(x_{tx}, x_{rx}) e^{-jk_{x_{tx}}x_{tx}} e^{-jk_{x_{rx}}x_{rx}} dx_{tx} dx_{rx}. \end{aligned} \tag{4.9}$$

Substituting the expression for $\boldsymbol{s}(x_{tx}, x_{rx})$ from (4.1), and changing the order of integration yields,

$$\begin{aligned} & S(k_{x_{tx}}, k_{x_{rx}}) \quad = \\ & \iint\limits_{z'x'} \boldsymbol{\gamma}(x', z') \tilde{\xi}(k_{x_{tx}}, k_{x_{rx}}, x', z') dx' dz', \end{aligned} \tag{4.10}$$

where $\tilde{\xi}(k_{x_{tx}}, k_{x_{rx}}, x', z') = \mathrm{FT}_{2D}\{\xi(x_{tx}, x_{rx}, x', z')\}$ denotes the space-variant transfer function of the system. The 2D FT operator can be decomposed into two 1D Fourier Transforms. Namely,

$$\begin{aligned} & \tilde{\xi}(k_{x_{tx}}, k_{x_{rx}}, x', z') = \\ & \quad \mathrm{FT}_{1D}\{e^{-jkR(x_{tx}, x'; z')}\} \mathrm{FT}_{1D}\{e^{-jkR(x_{rx}, x'; z')}\}, \end{aligned} \tag{4.11}$$

where $\mathrm{FT}_{1D}\{f(\alpha)\} \triangleq \int_\alpha f(\alpha)e^{-jk_\alpha\alpha}d\alpha$. By using the method of stationary phase [114, 115], which provides an approximate solution for integrals of oscillatory functions, we can compute the 1D Fourier Transforms in (4.11). The method is based on evaluating the positions of the function where the phase is stationary (i.e., the derivative is equal to zero), and replacing the integral with the sum of the function values at the stationary points. It is easy to see that the stationary points x_{tx}^{sp} and x_{rx}^{sp} corresponding to $\mathrm{FT}_{1D}\{e^{-jkR(x_{tx},x';z')}\}$ and $\mathrm{FT}_{1D}\{e^{-jkR(x_{rx},x';z')}\}$, respectively, satisfy the following,

$$\begin{aligned} k_{x_{tx}}R(x_{tx}^{sp},x';z') &= k(x'-x_{tx}^{sp}) \\ k_{x_{rx}}R(x_{rx}^{sp},x';z') &= k(x'-x_{rx}^{sp}). \end{aligned} \tag{4.12}$$

Substituting x_{tx}^{sp} and x_{rx}^{sp} in the FT arguments, and performing some algebraic manipulations yields,

$$\begin{aligned} \mathrm{FT}_{1D}\{e^{-jkR(x_{tx},x';z')}\} &\approx e^{-jk_{z_{tx}}(z'-z_a)}e^{-jk_{x_{tx}}x'} \\ \mathrm{FT}_{1D}\{e^{-jkR(x_{rx},x';z')}\} &\approx e^{-jk_{z_{rx}}(z'-z_a)}e^{-jk_{x_{rx}}x'}, \end{aligned} \tag{4.13}$$

where $k_{z_{tx}} = \sqrt{k^2-k_{x_{tx}}^2}$ and $k_{z_{rx}} = \sqrt{k^2-k_{x_{rx}}^2}$ are the spatial frequencies in the z direction for the Tx and Rx, respectively. Substituting (4.11) and (4.13) in (4.10) yields,

$$\begin{aligned} &S(k_{x_{tx}},k_{x_{rx}}) \approx e^{j(k_{z_{tx}}+k_{z_{rx}})z_a} \times \\ &\iint_{z'x'} \boldsymbol{\gamma}(x',z')e^{-j(k_{x_{tx}}+k_{x_{rx}})x'}e^{-j(k_{z_{tx}}+k_{z_{rx}})z'}dx'dz' \\ &= e^{j(k_{z_{tx}}+k_{z_{rx}})z_a}\tilde{\boldsymbol{\gamma}}(k_{x_{tx}}+k_{x_{rx}},k_{z_{tx}}+k_{z_{rx}}), \end{aligned} \tag{4.14}$$

where $\tilde{\boldsymbol{\gamma}}(k_x,k_z) \triangleq \mathrm{FT}_{2D}\{\boldsymbol{\gamma}(x',z')\}$, is the 2D spectrum of the scene reflectivity function. Let us define the 2D k-space vectors corresponding to the Tx and Rx as $\mathbf{k}_{tx} = (k_{x_{tx}},k_{z_{tx}})$,

and $\mathbf{k}_{rx} = (k_{x_{rx}}, k_{z_{rx}})$, respectively. From equation (4.14), we see that the spectrum of the scene has been sampled at,

$$\begin{aligned} \mathbf{k}_{\boldsymbol{\gamma}} \triangleq (k_x, k_z) &= \mathbf{k}_{tx} + \mathbf{k}_{rx} \\ &= (k_{x_{tx}} + k_{x_{rx}}, k_{z_{tx}} + k_{z_{rx}}), \end{aligned} \tag{4.15}$$

i.e., $\mathbf{k}_{\boldsymbol{\gamma}}$ is the summation of two vectors $\mathbf{k}_{tx}$ and $\mathbf{k}_{rx}$, each of which lie on the ring of radius k.

Viewing angles for point scatterers: The method of stationary phase characterizes the spatial frequency components $\mathbf{k}_{tx}$ and $\mathbf{k}_{rx}$ corresponding to the dominant propagating plane waves, for a given point scatterer located at (x', z'), and a given bistatic Tx/Rx pair (depicted in Figure 4.2-a). Stationary phase conditions in (4.12), describe the dominant spatial frequency components as

$$\begin{aligned} k_{x_{tx}} &= k\sin(\theta_{tx}), \quad k_{x_{rx}} = k\sin(\theta_{rx}), \\ k_{z_{tx}} &= k\cos(\theta_{tx}), \quad k_{z_{rx}} = k\cos(\theta_{rx}). \end{aligned} \tag{4.16}$$

where θ_{tx} and θ_{rx} denote the viewing angles shown in Figure 4.2-a. Therefore, using (4.15) and (4.16) we can characterize the k-space spectrum of a point scatterer for any given array architecture. For instance, Figure 4.2-b shows the sampled point ($\mathbf{k}_{\boldsymbol{\gamma}}$) in the spectrum of the point scatterer in Figure 4.2-a. See [115, 102] for more details on the k-space representation of active imaging systems.

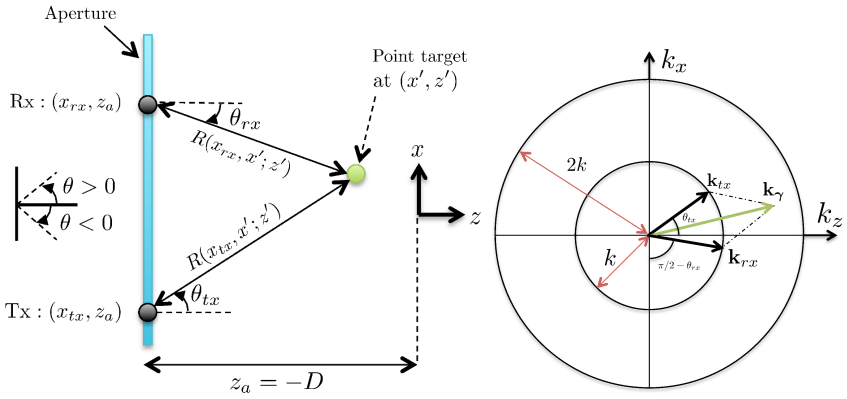


Figure 4.2: (left) The geometry of bistatic Tx/Rx elements and a point scatterer in the scene, (right) the corresponding sampled point in the spectrum of the point scatterer in k-space.

4.3 k-space Spectrum Corresponding to Different Array Architectures

Consider the 1-dimensional array geometry depicted in Figure 4.3. Let us restrict $\boldsymbol{\gamma}(x', z')$ to a plane parallel to the aperture with reflectivity $\boldsymbol{\gamma}(x') \triangleq \boldsymbol{\gamma}(x', z' = 0)$. Let L_1, L_2, and D denote the size of the aperture, the size of the scene, and the distance between the aperture and the scene, respectively. For any point-scatterer located at $x = x'$ we can identify the spectral region that will be sampled using a specific array geometry. A monostatic architecture restricts the Tx and Rx to be co-located, hence, for any array element we have $\mathbf{k}_{tx} = \mathbf{k}_{rx}$. By equation (4.15), we have $\mathbf{k}_{\boldsymbol{\gamma}} = 2\mathbf{k}_{tx}$, i.e., the ℓ_2-norm of $\mathbf{k}_{\boldsymbol{\gamma}}$ is $||\mathbf{k}_{\boldsymbol{\gamma}}||_2 = 2k$, and its direction is determined by the corresponding viewing angle $\theta_{tx} = \theta_{rx}$. Figure 4.4-a shows the spectral content (corresponding to the point-scatterer in Figure 4.3) seen through a monostatic array of infinitely many TRx elements. We see that $\mathbf{k}_{\boldsymbol{\gamma}}$ lies on the arc of the circle of radius $2k$, confined by the angles α and β, i.e., the two extremes of viewing angles from the aperture (also depicted in Figure 4.3).

Consequently, the monostatic spectrum corresponding to a point scatterer is given by

$$\mathcal{T}_{\text{mono}} = \{2\mathbf{k}_{tx} : ||\mathbf{k}_{tx}||_2 = k, \alpha \leq \angle \mathbf{k}_{tx} \leq \beta\}. \tag{4.17}$$

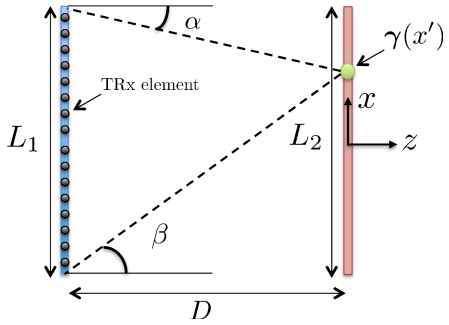


Figure 4.3: Geometry G1: 1D parallel and symmetric propagation model.

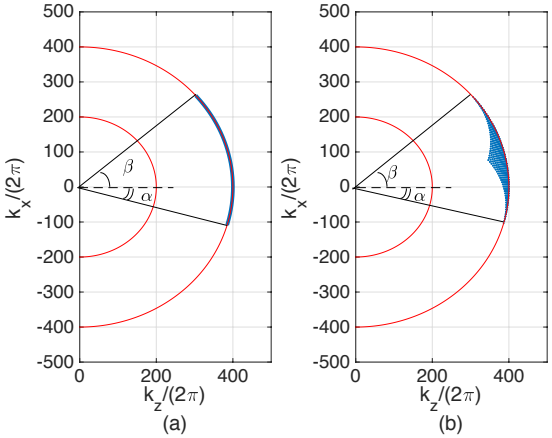


Figure 4.4: k-space spectrum corresponding to point-scatterer in Figure 4.3, seen through (left) monostatic and (right) multistatic array of infinitely many TRx elements.

On the other hand, in a multistatic array the Tx and Rx elements are not forced to be co-located, hence they can have different viewing angles. Figure 4.4-b shows the spectral content of the point-scatterer seen through a multistatic array. For $\theta_{tx} \neq \theta_{rx}$, we have

$$||\mathbf{k}_{\boldsymbol{\gamma}}||_2 = 2k\cos(|\theta_{tx} - \theta_{rx}|/2) < 2k.$$

Hence, multistatic array not only samples the region of the spectrum seen by a monostatic array, but also is able to sample points that are inside the circle of radius $2k$. Note that the *angular extent* of the sampled region determined by the extreme viewing angles α and β, is the same for monostatic and multistatic arrays. The multistatic spectrum corresponding to a point scatterer is given by

$$\begin{aligned} \mathcal{T}_{\text{multi}} &= \{\mathbf{k}_{tx} + \mathbf{k}_{rx} : ||\mathbf{k}_{tx}||_2 = ||\mathbf{k}_{rx}||_2 = k, \\ &\quad \alpha \leq \angle\mathbf{k}_{tx} \leq \beta, \alpha \leq \angle\mathbf{k}_{rx} \leq \beta\}. \end{aligned} \tag{4.18}$$

It is easy to see that $\mathcal{T}_{\text{multi}}$ can also be written as,

$$\mathcal{T}_{\text{multi}} = \{(p_1 + p_2)/2 : p_1, p_2 \in \mathcal{T}_{\text{mono}}\}. \tag{4.19}$$

Therefore,

$$\mathcal{T}_{\text{mono}} \subseteq \mathcal{T}_{\text{multi}} \subseteq \text{conv}(\mathcal{T}_{\text{mono}}), \tag{4.20}$$

where $\text{conv}(\mathcal{T})$ denotes the convex hull of the set $\mathcal{T}$.

Effective monostatic: As shown in Figure 4.2, a spatially-separated pair of Tx/Rx elements sample the spectrum of a point scatterer at $\mathbf{k}_{\boldsymbol{\gamma}}$, where $||\mathbf{k}_{\boldsymbol{\gamma}}|| = 2k\cos(|\theta_{tx} - \theta_{rx}|/2)$, and $\angle\mathbf{k}_{\boldsymbol{\gamma}} = (\theta_{tx} + \theta_{rx})/2$. Therefore, for a given point-scatterer in the scene, the same information can be captured by replacing the Tx/Rx pair with a monostatic element located at $x_{eff} \in [x_{tx}, x_{rx}]$, such that $\theta_{eff} = (\theta_{tx} + \theta_{rx})/2$, and transmitting a sinusoidal wave of wavelength $\lambda_{eff} = \lambda/\cos(|\theta_{tx} - \theta_{rx}|/2)$. Note that x_{eff} and λ_{eff} depend on the viewing angles of the point-scatterer, and can not be generalized to the entire scene. In Section 4.5 we investigate Fresnel approximation (in the far field) to overcome this barrier and define an effective monostatic array that applies to the entire scene.

4.4 Space-Bandwidth Product and The Degrees of Freedom

We are interested in identifying the number of DoF, i.e., the number of independent parameters that is possible to extract from an arbitrary scene, assuming that we are only constrained by the geometry of the imaging scenario. In this section, we introduce *space-bandwidth product* (SBP), defined by the product of the scene area and the measured spectral extent of the scene (after removing redundancies), as a means of identifying the DoF of the system. SBP can be thought of as a generalization of the so-called Shannon number [31], for spectral measurements of an unknown scene through a space-variant bandlimitted system. We evaluate the accuracy of SBP measure by comparing it to numerical SVD computations for various geometries.

The scene information lies in $\boldsymbol{\gamma}(x')$. We do not assume any constraints on the reflectivity function, namely it can take any complex value for each position x'. Substituting $\boldsymbol{\gamma}(x') = \boldsymbol{\gamma}(x', z' = 0)$ in (4.14), gives

$$\begin{aligned} S(k_{x_{tx}}, k_{x_{rx}}) &= e^{j(k_{z_{tx}}+k_{z_{rx}})z_a} \int_{x'} \boldsymbol{\gamma}(x') e^{-j(k_{x_{tx}}+k_{x_{rx}})x'} dx' \\ &= e^{jk_z z_a} \int_{x'} \boldsymbol{\gamma}(x') e^{-jk_x x'} dx'. \end{aligned} \tag{4.21}$$

Note that the integral kernel in (4.21) only depends on $k_x = k_{x_{tx}} + k_{x_{rx}}$, i.e., any pair of Tx/Rx elements that lead to the same spatial frequency in the x direction, capture the exact same information about the scene. Hence, in order to avoid redundancy in the acquired information, we consider the *projection* of the sampled points in the spectrum onto the k_x axis. For any point-scatterer located at $x = x'$, let us define the *spatial frequency bandwidth* $B(x')$ as the width of the corresponding spectrum after the projection

operation. Figure 4.5 shows $B(x')$ corresponding to the point-scatterer in Figure 4.3, for monostatic and multistatic array architectures. The following theorem will be useful in characterizing the SBP of 1D imaging system.

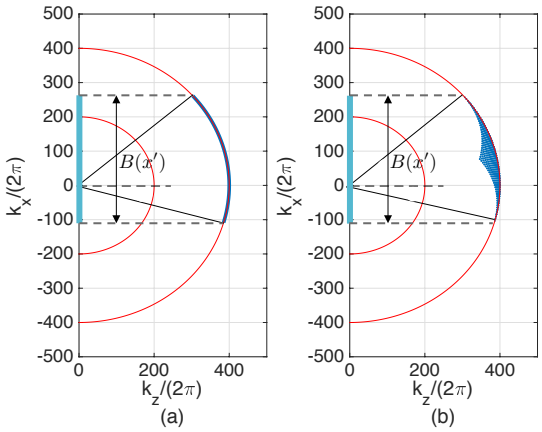


Figure 4.5: Spatial frequency bandwidth $B(x')$ corresponding to point-scatterer in Figure 4.3, computed after projection of the spectrum onto k_x axis for (left) monostatic and (right) multistatic array of infinitely many TRx elements.

Theorem 4 *Let $\hat{\mathcal{T}} = \mathcal{I}_l(\mathcal{T})$ denote the mapping operation, projecting set $\mathcal{T}$ onto the line l in 2D space. Then,*

$$\mathcal{I}_l(\mathcal{T}_{mono}) = \mathcal{I}_l(\mathcal{T}_{multi}), \ \forall l, \tag{4.22}$$

for any point scatterer in the scene. Consequently, the spatial frequency bandwidth $B(x')$, is the same for monostatic and multistatic architectures. See Appendix C for the proof.

The rest of this section is devoted to analytic and numerical computation of SBP for different imaging geometries. We consider the nominal values for wavelength $\lambda = 0.5$cm (corresponding to 60GHz temporal frequency), the size of the aperture $L_1 = 15$cm, the size of the scene $L_2 = 10$cm, and the distance between the aperture and the scene $D = 20$cm for the simulations in Section 4.4, unless stated otherwise.

4.4.1 SBP for 1D parallel planes geometry

Consider the parallel planes geometry G2 with translation parameter t, shown in Figure 4.6. For a point scatterer located at $x = x'$, we have

$$B(x') = \frac{2}{\lambda}\left(\sin(\beta(x')) - \sin(\alpha(x'))\right), \tag{4.23}$$

where,

$$\sin(\beta(x')) = \frac{(x' - a_1)}{\sqrt{D^2 + (x' - a_1)^2}} \tag{4.24}$$

$$\sin(\alpha(x')) = \frac{(x' - a_2)}{\sqrt{D^2 + (x' - a_2)^2}}. \tag{4.25}$$

For a small segment of the scene of length dx', the SBP is approximately given by $B(x')dx'$. Therefore, the total SBP of the scene, is calculated by the integral

$$\text{SBP} = \int_{\text{scene}} B(x')dx'. \tag{4.26}$$

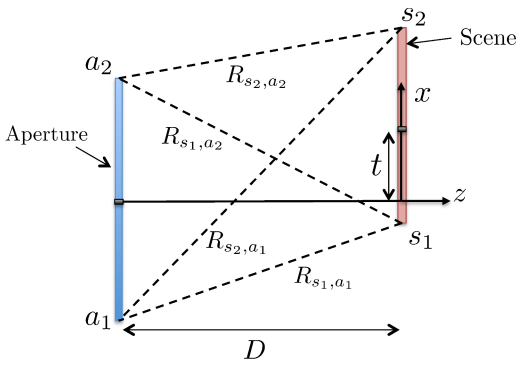


Figure 4.6: Geometry G2: 1D parallel planes propagation model, with an arbitrary scene translation t.

Assuming the scene reflectivity function $\boldsymbol{\gamma}(x') = 0$ for $x' \notin [s_1, s_2]$, and the aperture spanning an interval $[a_1, a_2]$ (as shown in Figure 4.6), SBP is calculated by

$$\begin{aligned} \text{SBP}_{G2} &= \int_{s_1}^{s_2} B(x')dx' \\ &= \int_{s_1}^{s_2} \frac{2}{\lambda}(\sin(\beta(x')) - \sin(\alpha(x')))dx' \\ &= \frac{2}{\lambda}\Big((R_{s_2,a_1} - R_{s_2,a_2}) + (R_{s_1,a_2} - R_{s_1,a_1})\Big) \end{aligned} \tag{4.27}$$

where $R_{i,j}$ denotes the distances between points i and j. Figure 4.7 shows the singular values of the normalized integral operator for $t = 15$cm for monostatic and multistatic architectures. We see that SBP_{G2}, shown by the dashed lines, is accurately predicting the critical point after which the singular values drop quickly to zero. We have conducted simulations for various values of parameters L_1, L_2, D and t, to verify the accuracy of SBP_{G2} in identifying the DoF for geometry G2. Figure 4.8 shows the variation of SBP_{G2} as a function of the translation parameter t, at various distances. We see that the maximum SBP_{G2} is obtained at $t = 0$ for all D, with the sensitivity of SBP to t being inversely related to D.

For the special case of symmetric and parallel geometry G1 (shown in Figure 4.3), we have $[a_1, a_2] = [-L_1/2, L_1/2]$ and $[s_1, s_2] = [-L_2/2, L_2/2]$, hence, the space-bandwidth product is given by

$$\begin{aligned} \text{SBP}_{G1} &= \frac{4}{\lambda}(R_{s_2,a_1} - R_{s_2,a_2}) \\ &= \frac{4D}{\lambda}\left(\sqrt{1 + \left(\frac{L_1 + L_2}{2D}\right)^2} - \sqrt{1 + \left(\frac{L_1 - L_2}{2D}\right)^2}\right). \end{aligned} \tag{4.28}$$

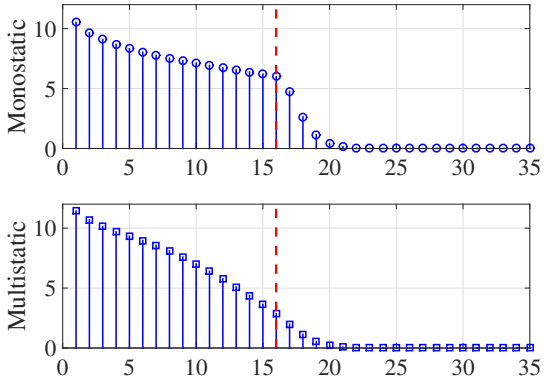


Figure 4.7: SVD analysis for parallel asymmetric geometry G2, with $t = 15$cm for (up) monostatic and (down) multistatic array of $N = 200$ array elements. Note that for this geometry $\text{SBP}_{G2} \approx 16$, depicted by the dashed line.

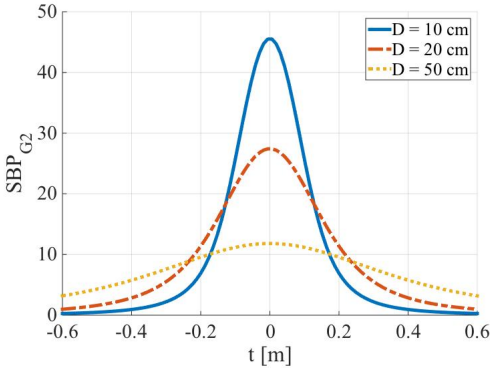


Figure 4.8: SBP computed for geometry G2 as a function of scene translation parameter t, with $L_1 = 15$ cm, $L_2 = 10$ cm, and $D \in \{10, 20, 50\}$ cm.

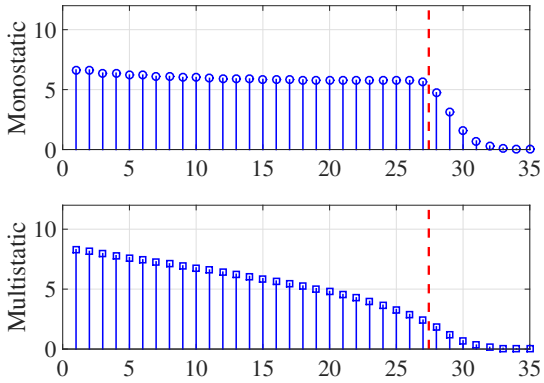


Figure 4.9: SVD analysis for nominal symmetric geometry G1, for (up) monostatic and (down) multistatic array of $N = 200$ array elements. Note that for this geometry $\mathrm{SBP}_{G1} \approx 27.4$, depicted by the dashed line.

For unbounded aperture $\lim_{L_1 \to \infty} \mathrm{SBP}_{G1} = \frac{4L_2}{\lambda}$, (similarly, for unbounded scene we have $\lim_{L_2 \to \infty} \mathrm{SBP}_{G1} = \frac{4L_1}{\lambda}$), i.e., SBP_{G1} does not increase indefinitely with increasing the size of the aperture (or the size of the scene). We note that our SBP_{G1} calculations are consistent with the heuristics reported in [103], and the explicit derivations in [105] for bounded and unbounded observation domains. As depicted in Figure 4.9, SBP_{G1} can accurately predict the number of DoF for the nominal geometry G1. Figure 4.10 summarizes the behavior of SBP_{G1} as a function of D and L_2. In particular Figure 4.10-b shows that SBP_{G1} reaches a plateau as L_2 is increased, with the corresponding upperbound being independent of D.

4.4.2 SBP for 1D rotated planes geometry

Consider the geometry G3 depicted in Figure 4.11, where the scene creates an angle θ with the x coordinate. In this scenario, the reflectivity function $\boldsymbol{\gamma}(x', z')$ is restricted to the line $x' = \rho z'$, where $\rho = \frac{-1}{\tan(\theta)}$, i.e., $\boldsymbol{\gamma}(x', z') = 0$ for all $x' \neq \rho z'$. Rewriting equation

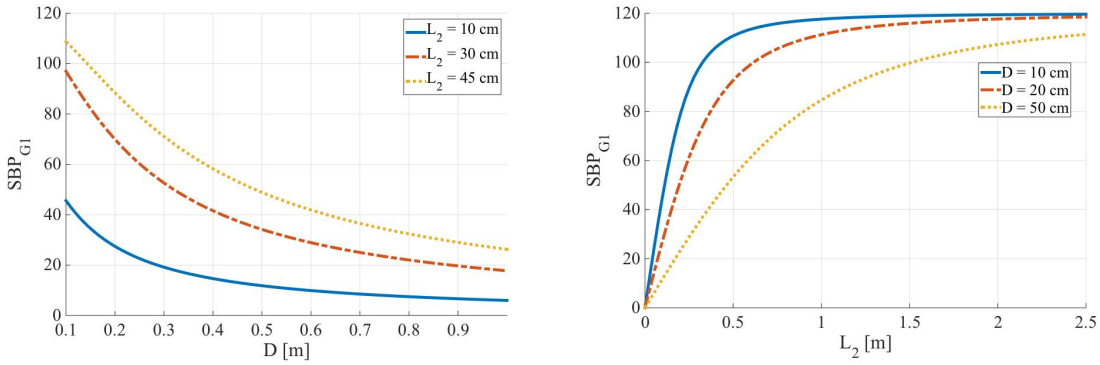


Figure 4.10: SBP computed for geometry G1 with $L_1 = 15$ cm fixed, as a function of (left) distance D, and (right) scene extent L_2.

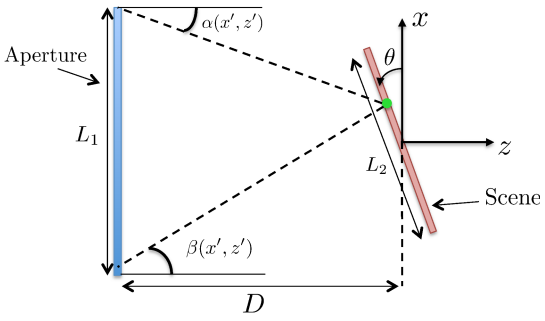


Figure 4.11: Geometry G3: 1D rotated planes propagation model.

(4.14) under this constraint leads to

$$\begin{aligned} S(k_{x_{tx}}, k_{x_{rx}}) &= e^{j(k_{z_{tx}}+k_{z_{rx}})z_a} \times \\ &\int_{z'} \boldsymbol{\gamma}(\rho z', z') e^{-j(k_{x_{tx}}+k_{x_{rx}})\rho z'} e^{-j(k_{z_{tx}}+k_{z_{rx}})z'} dz' \\ &= \int_{z'} \boldsymbol{\gamma}(\rho z', z') e^{-j(\rho k_x + k_z)z'} dz'. \end{aligned} \tag{4.29}$$

The integral kernel in (4.29) depends on $\rho k_x + k_z$, i.e., any pair of Tx/Rx elements that lead to the same value of $\rho k_x + k_z$, deliver the same information about the reflectivity function. In order to avoid redundancy in the acquired information through the imaging system, we need to project the sampled points in the spectrum of the scene onto the line $k_x = \rho k_z$, i.e., the line that crosses the origin and creates an angle θ with the k_x coordinate, as shown in Figure 4.12. The space bandwidth product for this geometry is given by

$$\text{SBP}_{G3} = \int_{\text{scene}} B(x', z') d\mu(x', z'). \tag{4.30}$$

where $\mu(\cdot, \cdot)$ is the standard Lebesgue measure on $\mathcal{A}$. As shown in Theorem 1, for any angle θ, the spatial frequency bandwidth $B(x', z')$, and consequently SBP_{G3}, is the same for monostatic and multistatic arrays. This can be verified through numerical SVD computations shown in Figure 4.13 for $\theta = 35°$.

Figure 4.14 shows the result of numerical computation of SBP_{G3} for different rotation angles and at different ranges. The space-bandwidth product is maximized for $\theta = 0$, which corresponds to parallel symmetric geometry G1. Increasing θ leads to a decrease in SBP_{G3}, till a global minimum is achieved at $\theta = 90°$, i.e., when the planes are orthogonal to each other. Note that even for orthogonal planes geometry SBP_{G3} is bounded away from zero, with its value increasing as we decrease the distance D. This is indeed the reason behind the improvement in range resolution for continuous-wave imaging systems

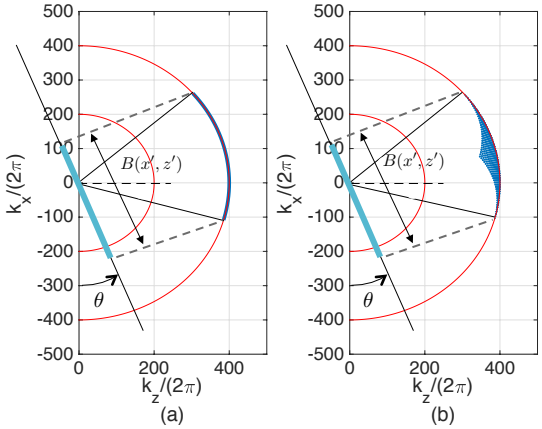


Figure 4.12: Spatial-frequency bandwidth corresponding to the point scatterer located at (x', z') for (left) monostatic and (right) multistatic arrays.

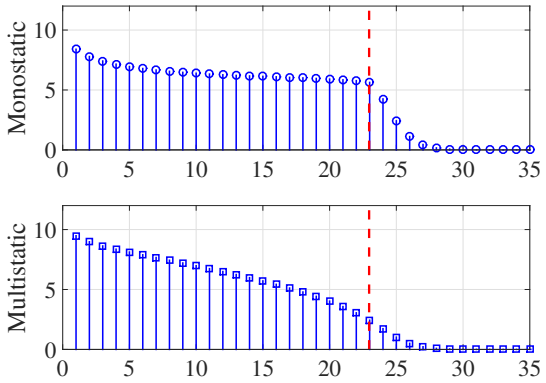


Figure 4.13: SVD analysis for rotated geometry G3, with $\theta = 35°$ for (up) monostatic and (down) multistatic array of $N = 200$ array elements. Note that for this geometry $\text{SBP}_{G3} \approx 23$, depicted by the dashed line.

at short range [115].

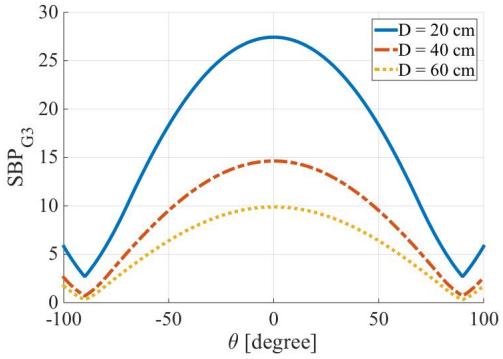


Figure 4.14: SBP computed for geometry G3 as a function of scene rotation θ, with $L_1 = 15$ cm, $L_2 = 10$ cm, and $D \in \{20, 40, 60\}$ cm.

4.4.3 SBP for 1D rotated and translated planes geometry

Let us consider the geometry shown in Figure 4.15, where the scene is rotated and translated simultaneously. The scene reflectivity function is restricted to the line $x' = \rho z' + t$, where $\rho = \frac{-1}{\tan(\theta)}$. Following similar arguments as in Subsection 4.4.2, one can show that the spatial frequency bandwidth for any point scatterer in the scene is evaluated by projecting the sampled points in the scene spectrum onto the line $k_x = \rho k_z$ (as depicted in Figure 4.12). We compute SBP_{G4} numerically for different realizations of the scene parameters. Figure 4.16 shows the singular values and the corresponding SBP_{G4} for $\theta = 55^\circ$ and $t = 20$cm. The variation of SBP_{G4} as a function of θ for multiple values of t is depicted in Figure 4.17. It is interesting to determine $\theta_{max}(t)$, i.e., the rotation angle that maximizes SBP_{G4} for a given t. One heuristic approach is based on choosing θ such that the scene is orthogonal to the line that connects the midpoints of the aperture and the scene, i.e., $\theta_{heu}(t) = \sin^{-1}(t/\sqrt{t^2 + D^2})$. Figure 4.18 shows that at short distances θ_{heu} overestimates θ_{max} for $t > 0$, with the difference vanishing as D is increased.

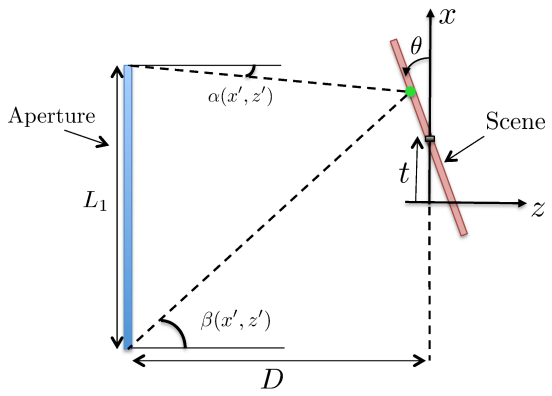


Figure 4.15: Geometry G4: 1D rotated and translated planes propagation model.

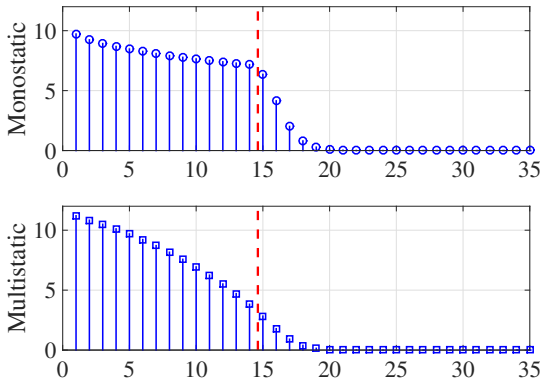


Figure 4.16: SVD analysis for geometry G4 with $t = 20$ cm, and $\theta = 55°$, for (up) monostatic and (down) multistatic array of $N = 200$ array elements. Note that for this geometry $\text{SBP}_{G4} \approx 14.6$, depicted by the dashed line.

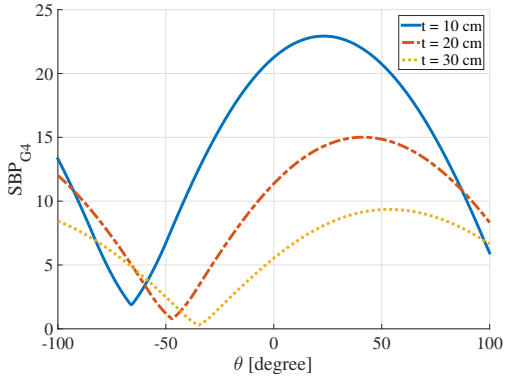


Figure 4.17: SBP computed for geometry G4, with $L_1 = 15$cm, $L_2 = 10$cm, $D = 20$cm, and $t \in \{10, 20, 30\}$ cm.

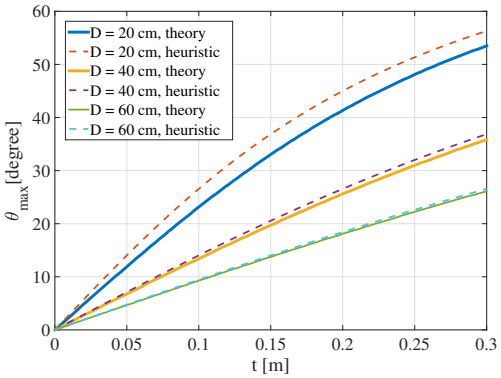


Figure 4.18: Rotation angle corresponding to the maximum SBP_{G4} as a function of translation parameter t, computed for $D \in \{20, 40, 60\}$ cm. The dashed and solid curves correspond to θ_{heu} and true θ_{max}, respectively.

4.5 Fresnel Approximation and Asymptotic Regime

In this section, we study monostatic and multistatic arrays under Fresnel approximation [102], that is we use first order Taylor approximation for computing path lengths in (4.1) when $D \gg L_1, L_2$. For the geometry G1 shown in Figure 4.3, Fresnel approximation yields

$$R(x, x'; z' = 0) = \sqrt{(x - x')^{+} D^2} \approx D + \frac{(x - x')^2}{2D}. \tag{4.31}$$

Therefore, Fresnel diffraction integral is given by

$$\begin{aligned} s(x_{tx}, x_{rx}) &= \int_{x'} \boldsymbol{\gamma}(x') e^{-jkR(x_{tx}, x'; 0)} e^{-jkR(x_{rx}, x'; 0)} dx' \\ &\approx e^{-j2kD} \int_{x'} \boldsymbol{\gamma}(x') e^{-j\frac{k}{2D}(x_{tx} - x')^2} e^{-j\frac{k}{2D}(x_{rx} - x')^2} dx'. \end{aligned} \tag{4.32}$$

We investigate the implications of Fresnel approximation for array design and illustrate the connections between different array architectures in Fresnel zone.

4.5.1 Monostatic array in Fresnel zone

Monostatic architecture ensures that $x_{tx} = x_{rx}$, hence equation (4.32) reduces to

$$\begin{aligned} s(x_{tx}, x_{rx}) &\approx e^{-j2kD} \int_{x'} \boldsymbol{\gamma}(x') e^{-j\frac{k}{D}(x_{tx}-x')^2} dx' \\ &= e^{-j2kD} e^{-j\frac{k}{D}x_{tx}^2} \int_{x'} \boldsymbol{\gamma}(x') e^{-j\frac{k}{D}x'^2} e^{j(\frac{2kx_{tx}}{D})x'} dx' \\ &= e^{-j2kD} e^{-j\frac{k}{D}x_{tx}^2} \mathrm{FT}_{1D}\left\{\boldsymbol{\gamma}(x') e^{-j\frac{k}{D}x'^2}\right\}_{f \triangleq \frac{kx_{tx}}{\pi D}}, \end{aligned} \tag{4.33}$$

where we have used a change of variable $f \triangleq \frac{kx_{tx}}{\pi D}$ in computing the 1D Fourier transform. The quadratic-phase terms (i.e., $e^{-j\frac{k}{D}x'^2}$ and $e^{-j\frac{k}{D}x_{tx}^2}$) in (4.33) are known as *Fresnel phase mask* [102]. Multiplying the reflectivity function by this mask does not lead to any information loss of the original function, and its effect can be simply inverted using the complex conjugate mask. Therefore, the only information bottleneck in equation (4.33) is due to the Fourier Trasform operation. As we mentioned in Section 3.2, the Fourier kernel has been studied in detail by Slepian *et al* in the context of time-limited and band-limited functions [61]. It has been shown that the eigenfunctions of this integral equation are PSWFs, and the corresponding eigenvalues have the interesting property that they remain approximately equal until a critical transition point where they rapidly decay to zero. This transition point for the class of time-limited and band-limited signal is determined by the time-bandwidth product [101]. The equivalent of time-bandwidth product in (4.33) corresponds to,

$$\mathrm{DoF} = \Delta x' \Delta f = L_2(\frac{k\Delta x_{tx}}{\pi D}) = \frac{2L_1L_2}{\lambda D}, \tag{4.34}$$

where $\Delta x' = L_2$ and $\Delta x_{tx} = L_1$ are identified by the scene and aperture extent, respectively.

In Section 4.4.1, we derived a closed-form expression for the space-bandwidth product for the symmetric parallel planes geometry G1, without any assumption on the distance of the scene from the aperture. In the special case of $D \gg L_1, L_2$, we can use Fresnel approximation to evaluate SBP_{G1} as

$$\begin{aligned}\mathrm{SBP}_{G1} &= \frac{4}{\lambda}(R_{s_2,a_1} - R_{s_2,a_2}) \\ &\approx \frac{4D}{\lambda}\left(1 + \frac{1}{2}\left(\frac{L_1+L_2}{2D}\right)^2 - 1 - \frac{1}{2}\left(\frac{L_1-L_2}{2D}\right)^2\right) \\ &= \frac{2L_1L_2}{\lambda D}. \end{aligned} \tag{4.35}$$

This result is in agreement with the classical DoF analysis derived in (4.34). An alternative interpretation is by approximating the spatial frequency bandwidth $B(x') \approx B(0) = 2(\frac{2}{\lambda})\sin(\alpha(0)) \approx \frac{4}{\lambda}\frac{L_1}{2D} = \frac{2L_1}{\lambda D}$ for all x', and hence, $\mathrm{SBP}_{G1} \approx L_2 B(0) = \frac{2L_1L_2}{\lambda D}$. This interpretation of the SBP calculation in Fresnel zone can be easily extended to geometry G3 (Figure 4.11) by projecting the spatial frequency bandwidth $B(0)$ onto the line that crosses the origin and creates an angle θ with the k_x coordinate, to obtain the following approximate formula,

$$\mathrm{SBP}_{G3} \approx \frac{2L_1L_2}{\lambda D}\cos(\theta). \tag{4.36}$$

It is worth mentioning that the seminal work of Slepian on PSWFs and the study of Fourier kernel has been applied in various engineering problems, in particular in the context of diffraction limited optics [91], and line-of-sight MIMO communications [27].

4.5.2 Multistatic array in Fresnel zone

In this subsection, we investigate multistatic imaging arrays under Fresnel approximation. The Fresnel diffraction integral in (4.32) for an arbitrary Tx/Rx pair can be

further simplified as,

$$s(x_{tx}, x_{rx}) \approx e^{-j2kD} e^{-j\frac{k}{4D}(x_{tx}-x_{rx})^2} \times \int_{x'} \boldsymbol{\gamma}(x') e^{-j\frac{k}{D}(x'-x_{mid})^2} dx', \tag{4.37}$$

where $x_{mid} \triangleq \frac{1}{2}(x_{tx} + x_{rx})$ represents the *midpoint* of the Tx/Rx pair. Comparing (4.37) with (4.33), we can see that the information being captured by the multistatic Tx/Rx pair is equivalent to that of a monostatic transceiver located at x_{mid}. Following similar lines of reasoning as we did for the analysis of monostatic arrays in Subsection 4.5.1, we can identify the number of the degrees of freedom for the integral kernel in (4.37) by,

$$\text{DoF} = \Delta x' \Delta f = L_2(\frac{k\Delta x_{mid}}{\pi D}) = \frac{2L_1 L_2}{\lambda D}. \tag{4.38}$$

Fresnel zone analysis of 1D multistatic architecture through PSWFs theory has also appeared in [116]. In derivation of (4.38) we have used the observation that x_{mid} is restricted to the aperture of the imaging system, hence the extent of its admissible values is identified by $\Delta x_{mid} = L_1$. This result agrees with our previous observation through the space-bandwidth product analysis in Section 4.4, where we established that the SBPs achieved by monostatic and multistatic architectures are equal to each other.

The approximate integral equation in (4.37) has significant practical implications for multistatic array design. Most importantly, (4.37) implies that a multistatic architecture can be replaced with an *effective monostatic* array, by placing a monostatic transceiver element at the midpoints of every Tx/Rx pair. Let $a_{tx}(x)$ and $a_{rx}(x)$ denote aperture functions that encode the locations of the transmitter and receiver elements, respectively.

Namely,

$$a_{tx}(x) = \sum_{i=1}^{N_{tx}} \delta(x - x_{tx}(i)),$$
$$a_{rx}(x) = \sum_{j=1}^{N_{rx}} \delta(x - x_{rx}(j)), \tag{4.39}$$

where $\delta(x)$ is the Dirac delta function. By definition, the effective monostatic array is given by,

$$a_{eff}(x) \triangleq \sum_{i=1}^{N_{tx}} \sum_{j=1}^{N_{rx}} \delta\left(x - \left(\frac{x_{tx}(i) + x_{rx}(j)}{2}\right)\right). \tag{4.40}$$

As shown in Appendix D, $a_{eff}(x)$ can also be expressed as,

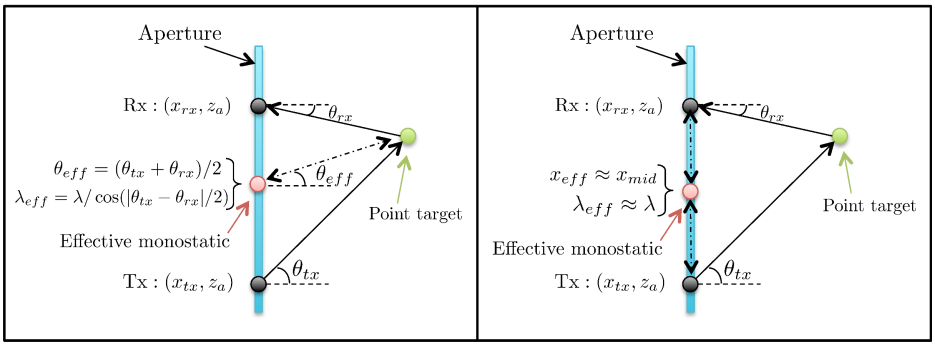
$$a_{eff}(x) = a_{tx}(2x) * a_{rx}(2x), \tag{4.41}$$

that is the effective monostatic array is derived by *shrinking* the Tx and Rx aperture functions by a factor of 2, followed by a *convolution* in the spatial domain. In the literature, the convolution expression for describing the effective monostatic array is mainly justified through *array factor* arguments (although the shrinkage step is missing in the array factor argument) [117, 118]. See Appendix D for a quick review of array factor arguments for the analysis of effective monostatic array.

In Section 4.3, we showed that for a given point-scatterer in the scene we can replace any spatially-separated Tx/Rx pair by a monostatic element that captures the exact same information. However, the constructed effective monostatic array is not generalizable to the entire scene and depends on the particular point-scatterer that has been considered. Here, we used Fresnel approximation to construct an effective monostatic that is independent from the scene and provides a systematic approach for designing and analyzing multistatic architectures in the Fresnel zone. Figure 4.19 summarizes the

construction of the effective monostatic array using Fresnel approximation, compared to the exact solution provided by the k-space analysis.

Remark 3 *Our SBP analysis reveals that multistatic array with an infinite number of array elements, does not lead to an improvement in the number of DoF compared to that of a monostatic array for 1D imaging scenario. However, from a practical point of view, it is desired to design a finite-element array with minimum number of transceivers that is capable of capturing the entire DoF. The effective array argument suggests that it is possible to realize a dense effective array, by intelligent co-design of sparse transmitter and receiver arrays. Moreover, in the presence of noise, deploying a multistatic architecture leads to an improvement in the signal to noise ratio (SNR) compared to a monostatic with the same number of elements.*



(a) (b)

Figure 4.19: Effective monostatic element corresponding to a bistatic pair (left) using k-space analysis and beyond Fresnel approximation, and (right) after Fresnel approximation.

In the next subsection, we take a more practical viewpoint and discuss the significance of the DoF argument in designing sparse monostatic arrays while avoiding grating lobes. We note that in our previous work [28], we presented a more detailed analysis of model-based sparse monostatic imaging systems, and here, we briefly review some of the key results and their connection to the DoF analysis.

4.6 Image Formation and Resolution Analysis

Image formation techniques aim at reconstructing the reflectivity function of the scene from the measured data, by solving the inverse scattering problem [114]. Inverse scattering has wide applications in medical imaging, non-destructive testing, optics, remote sensing, etc. [114, 119, 120]. In general, the inverse scattering problems are ill-posed due to the non-trivial nullspace of the imaging systems, and hence, the reflectivity function that satisfies our measurement equations is not unique. In this section, we study the implications of the DoF analysis for image reconstruction, and analyze the achievable resolution in the image.

4.6.1 Pseudoinverse reconstruction

Since the number of DoF of the imaging systems is finite, we can rewrite equation (4.5) as,

$$\boldsymbol{s} \approx \sum_{i=1}^{\text{DoF}} \sigma_i \phi_i \langle \boldsymbol{\gamma}, \psi_i \rangle_{\Psi}. \tag{4.42}$$

This regularization procedure is known as Truncated SVD (TSVD) [104, 112], where we only keep the non-zero singular values with a significant contribution to the measured data and set the rest of the singular values to zero. Assuming noiseless measurements, the "best approximate" solution to the image formation problem is given by the minimum ℓ^2-norm estimate of the scene reflectivity function consistent with the data. This solution is obtained by computing $\Xi^{\dagger}\boldsymbol{s}$, where $\Xi^{\dagger} : \Phi \to \Psi$ is the Moore-Penrose pseudoinverse (PINV) of Ξ. The reconstructed image is given by,

$$\hat{\boldsymbol{\gamma}}_{pinv} = \sum_{i=1}^{\text{DoF}} \frac{1}{\sigma_i} \psi_i \langle \boldsymbol{s}, \phi_i \rangle_{\Phi}. \tag{4.43}$$

Combining (4.42) and (4.43), gives us,

$$\hat{\boldsymbol{\gamma}}_{pinv} = \sum_{i=1}^{\mathrm{DoF}} \psi_i \langle \boldsymbol{\gamma}, \psi_i \rangle_{\Psi}, \tag{4.44}$$

i.e., the image is formed by projecting the original reflectivity function $\boldsymbol{\gamma}$ onto the subspace spanned by $\{\psi_i : i = 1, \ldots, \mathrm{DoF}\}$. Note that the reconstruction error $(\boldsymbol{\gamma} - \hat{\boldsymbol{\gamma}}_{pinv})$ lies approximately in the null-space of Ξ, i.e., $\Xi(\boldsymbol{\gamma} - \hat{\boldsymbol{\gamma}}_{pinv}) \approx 0$, hence not observed through the imaging system. See [121] for more details on TSVD-based pseudoinverse operation.

4.6.2 Matched-filter/back-propagation reconstruction

The standard and classical method for image reconstruction is based on applying the Hermitian adjoint operator to the measured data,

$$\hat{\boldsymbol{\gamma}}_{adj} = \Xi^{\ddagger} \boldsymbol{s}, \tag{4.45}$$

where $\Xi^{\ddagger} : \Phi \to \Psi$ denoted the adjoint of Ξ. This procedure is also known as *Matched-Filtering* (MF) [122], and *Back-Propagation* algorithm [123]. The integral operation corresponding to (4.45) is identified by,

$$\hat{\boldsymbol{\gamma}}_{adj}(x'', z'') = \int_{\mathcal{B}} \xi^*(x_{tx}, x_{rx}, x'', z'') \boldsymbol{s}(x_{tx}, x_{rx}) dx_{tx} dx_{rx}, \tag{4.46}$$

where ξ^* is the complex conjugate of ξ (It is easy to verify that $\langle \Xi\boldsymbol{\gamma}, \boldsymbol{s} \rangle_{\Phi} = \langle \boldsymbol{\gamma}, \Xi^{\ddagger}\boldsymbol{s} \rangle_{\Psi}$). Combining (4.1) and (4.46) gives us,

$$\hat{\boldsymbol{\gamma}}_{adj}(x'', z'') = \int_{\mathcal{A}} \kappa(x'', z'', x', z') \boldsymbol{\gamma}(x', z') dx' dz', \tag{4.47}$$

where κ corresponds to the compact self-adjoint linear operator defined by,

$$\kappa(x'', z'', x', z') = \int_{\mathcal{B}} \xi^*(x_{tx}, x_{rx}, x'', z'')\xi(x_{tx}, x_{rx}, x', z')dx_{tx}dx_{rx}. \tag{4.48}$$

Using (4.6), we can rewrite (4.47) as,

$$\hat{\boldsymbol{\gamma}}_{adj} = \sum_{i=1}^{\infty} \sigma_i^2 \psi_i \langle \boldsymbol{\gamma}, \psi_i \rangle_{\Psi} \approx \sum_{i=1}^{\text{DoF}} \sigma_i^2 \psi_i \langle \boldsymbol{\gamma}, \psi_i \rangle_{\Psi}, \tag{4.49}$$

i.e., projecting $\boldsymbol{\gamma}$ onto the subspace spanned by $\{\psi_i : i = 1, \ldots, \text{DoF}\}$, while *weighting* the components of the projection by the square of the corresponding singular values. Note that if the singular values of the imaging system are approximately equal, $\sigma_i \approx \sigma$, $\forall i$, then $\hat{\boldsymbol{\gamma}}_{adj} \approx \sigma^2 \hat{\boldsymbol{\gamma}}_{pinv}$, that is the image formed by the adjoint operator is just a scaled version of the output of Moore-Penrose pseudoinverse operator.

4.6.3 Resolution analysis

One of the most important performance metrics for any imaging systems is the resolution capability. Here we use the classical Rayleigh's criterion and the associated reciprocal bandwidth arguments [124], to compare the performance of different reconstructions schemes for monostatic and multistatic arrays. It follows from uncertainty principle that a function's width in the spatial domain is inversely proportional to its width in the spatial frequency domain [125], hence, we use $1/B(x', z')$ as a benchmark measure of achievable resolution for a point scatterer located at (x', z'). We fix $L_1 = 15$cm, $L_2 = 10$cm, and $D = 40$cm for the numerical results of this subsection.

Based on Rayleigh's criterion, the resolution of an imaging system is defined by the

spatial width of the reconstructed image corresponding to a point target, also known as point spread function (PSF). For a point target located at (x'_p, z'_p), we substitute $\boldsymbol{\gamma}(x', z') = \delta(x' - x'_p)\delta(z' - z'_p)$ in (4.44) and (4.49), to obtain PSFs corresponding to the pseudoinverse (Figure 4.20) and matched filter (Figure 4.21) reconstruction schemes, respectively. From (4.47), it is evident that $\kappa(x'', z'', x'_p, z'_p)$ also identifies the PSF corresponding to the matched filter scheme. In order to quantify the achievable resolution for different scenarios, we evaluate 3dB beamwidth of the mainlobe of the PSFs, and compare the results with $1/B(x', z')$. As depicted in Figure 4.22, for geometry G1, PINV reconstruction outperforms MF, and leads to a better resolution throughout the scene. More importantly, we see a significant resolution loss for MF method for multistatic architecture. This can be partially explained by our previous observation that if the variance of the singular values of the imaging system is large (which seems to be the case for multistatic arrays based on our SVD computations), then MF reconstruction deviates from the optimal PINV operation.

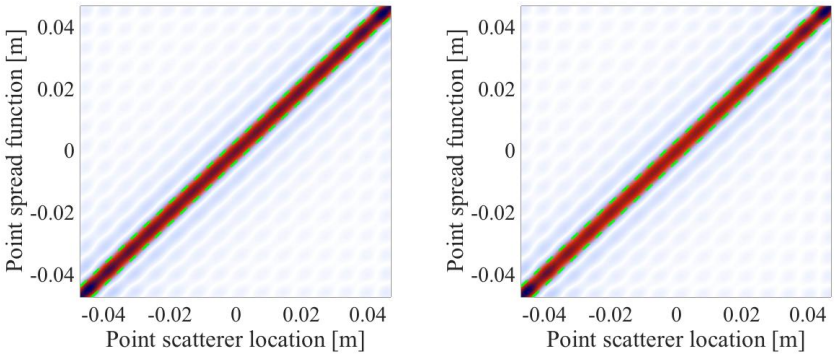


Figure 4.20: PSFs with PINV reconstruction corresponding to geometry G1, for (left) monostatic, and (right) multistatic arrays. The reciprocal spatial frequency bandwidth $1/B(x', z')$, is depicted by the dashed lines.

We also evaluate the achievable resolution corresponding to geometries G2 and G3, with $t = 15$cm and $\theta = 40°$, respectively. As depicted in Figure 4.23-a for geometry G2, the resolution is inversely related to the distance of the point scatterer from the center of

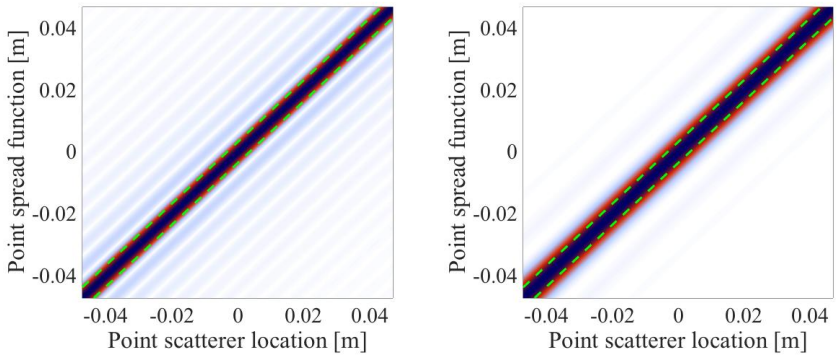


Figure 4.21: PSFs with MF reconstruction corresponding to geometry G1, for (left) monostatic, and (right) multistatic arrays. The reciprocal spatial frequency bandwidth $1/B(x', z')$, is depicted by the dashed lines.

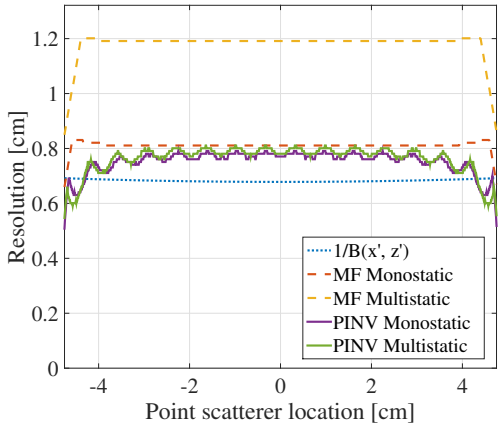


Figure 4.22: Resolution evaluated by 3dB beamwidth of the mainlobe of PSFs for geometry G1.

the aperture. Also, as shown in Figure 4.23-b for geometry G3, rotation of the scene leads to an improvement in resolution for the point scatterers that get closer to the aperture, and a degradation for the ones that get farther. These results are intuitively appealing as well as theoretically justifiable through reciprocal bandwidth arguments.

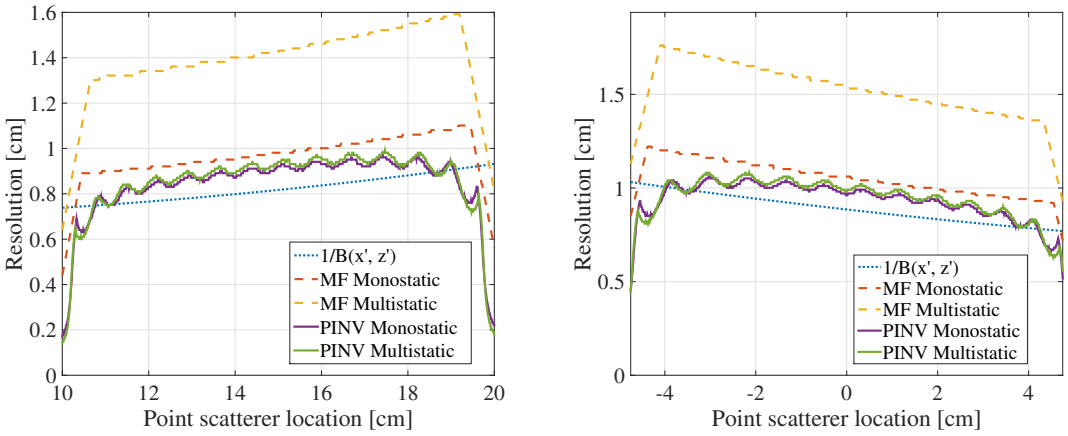


Figure 4.23: Resolution evaluated by 3dB beamwidth of the mainlobe of PSFs for (left) geometry G2 with $t = 15$cm, and (right) geometry G3 with $\theta = 40°$.

Remark 4 *We know that the number of DoF and the SVD (including the singular values and the singular functions) corresponding to an imaging scenario depend on the scene extent L_2, whereas the resolution of a point scatterer and its spatial frequency bandwidth $B(x', z')$ are independent of L_2. This leads us to an interesting observation: even though PINV reconstruction uses SVD for inversion (which is a global decomposition of the imaging scenario), its corresponding resolution limit is closely following the reciprocal bandwidth $1/B(x', z')$; a locally defined measure of accuracy in determining the location of the point scatterers.*

Chapter 5

Conclusions and Future Work

5.1 Near-Optimal Frequency Estimation

We have shown that NOMP is fast and near-optimal for frequency estimation in AWGN. It performs better than classical methods such as MUSIC, and more recent convex optimization based methods such as AST, in terms of both estimation accuracy and run time. The algorithm uses a fundamental element of OMP, ensuring that the residue is orthogonal to the signal space spanned by the current set of frequencies. However, NOMP avoids the error floors of naively discretized OMP by refinement over the continuum. Specifically, it searches for a signal subspace in the "neighborhood" of our current signal space which can better explain the observed measurements. The algorithm has a natural "decision feedback" interpretation in which it gives the already detected sinusoids a chance to adjust their frequencies in light of new evidence which is presented in the form of an updated residue after we add the next sinusoid.

We believe that the ideas underlying NOMP are broadly applicable to problems of sparse approximation over continuous dictionaries, but further work is required to justify this assertion. An open problem is to go beyond the pessimistic convergence rate

estimates provided here by analytically quantifying the benefits of refinement.

5.2 Sparsity-Driven Imaging

We have shown that short-range radar imaging requires new models that account for scene and transceiver geometry, as well as the number of transceiver elements. For sparse arrays, the patch-based models introduced here suppress grating lobes, while being compatible with the spatially lowpass nature of typical scenes. We have applied different reconstruction algorithms to demonstrate the resilience of patch-based models to the sparsity of the imaging array. It is worth mentioning that for the special case where we construct the dictionary of patches by spatial translations of a single patch (that is fixed in shape and size), patch-based MF is equivalent to applying a low-pass filter (LPF) in the spatial-frequency domain. We note that interesting fundamental insights into this problem can be obtained by pursuing spatial frequency domain analysis more deeply. This is a subject of ongoing work, and will appear in our future publications.

We have introduced a general framework for constructing spatially-extended dictionaries that provide a sparse representation for simple scenes, therefore encoding the information in a minimalistic manner. The sparsity can be further exploited for image reconstruction in order to super-resolve the scene beyond the capability of conventional methods. An important topic for future work, especially for sparse arrays, is to extract and employ Doppler information as well (see preliminary work by other members of our group in [126]).

5.3 Singular Decomposition and Degrees of Freedom

We have introduced the space-bandwidth product as a measure of predicting the number of DoF of 1D radar imaging systems under Born approximation (our results extend to 2D planar geometries. The analysis for 2D settings will appear in future publications). Our analysis goes beyond Fresnel approximation and provides insights for short-range scenarios, where Fresnel (and hence the results from PSWF theory) breaks down. We have validates the accuracy of SBP in predicting the DoF by numerical evaluations of the singular decomposition of the imaging system in various scenarios. Our preliminary results indicate that the singular decomposition approach may provide an analytical framework for sparse array design and analysis. For example: (i) for a typical scene, we can identify the singular functions that are most useful in encoding the scene information (e.g., the inner-product of the singular function and the scene is large), and construct spectral filters that are focused on capturing the relevant information in the scene, (ii) we can design sparse arrays based on a subset (or the entire set) of singular functions over the aperture, e.g., by optimizing the location of the array elements to capture a larger portion of the energy of the chosen singular functions, while trying to preserve the orthogonality property of the functions, (iii) more efficient image reconstruction techniques that are capable of incorporating prior information into the reconstruction procedure can be devised. Further investigation of these techniques is a subject of our future work, and will appear in forthcoming publications.

Appendix A

Bayesian Information Criteria

The BIC balances an increase in the likelihood with the number of parameters used to achieve that increase. Namely,

$$\text{BIC} = 2\ln\frac{L_2(\theta_1,\ldots,\theta_{m_2})}{L_1(\phi_1,\ldots,\phi_{m_1})} - (m_2 - m_1)\ln(N), \tag{A.1}$$

where $L_1(\phi_1,\ldots,\phi_{m_1})$ and $L_2(\theta_1,\ldots,\theta_{m_2})$ are likelihood functions and $\{\phi_i\}_{i=1}^{m_1}$ and $\{\theta_j\}_{j=1}^{m_2}$ are their corresponding parameters. For our measurement model, $\mathbf{y} = \sum_{l=1}^{K} g_l\mathbf{x}(\omega_l) + \mathbf{z}$, where $\mathbf{z} \sim \mathcal{N}(\mathbf{0}, \sigma^2\mathbb{I}_N)$, we have

$$L(\{g_l,\omega_l\}_{l=1}^{K}) = \frac{1}{(\pi\sigma^2)^N}\exp(-||\mathbf{y}_r||^2/\sigma^2),$$

where $\mathbf{y}_r = \mathbf{y} - \sum_{l=1}^{K} g_l\mathbf{x}(\omega_l)$ is the residual. Therefore, $\text{BIC} = \frac{2}{\sigma^2}\Delta||\mathbf{y}_r||^2 - 2\ln(N)$, where $\Delta||\mathbf{y}_r||^2 = ||\mathbf{y}_r(old)||^2 - ||\mathbf{y}_r(new)||^2$ is the reduction in residual energy after detecting a new sinusoid. If $\text{BIC} > 10$, it is a strong evidence that the most recent reduction in the residual energy corresponds to a newly detected sinusoid. Therefore, when $\text{BIC} < 10$, we stop the algorithm.

Appendix B

Cramér Rao Bound

In this Appendix we first review the Cramér Rao Bound [18] for an unbiased estimator in a general setting. Let $\mathbf{a} \in \mathbb{R}^{\lambda}$. If $\mathbf{a}^T\hat{\theta}(\mathbf{y})$ is an unbiased estimator of $\mathbf{a}^T\theta$, then the variance of the estimator given by $\mathbb{E}_{\mathbf{y}|\theta}\left[\left(\mathbf{a}^T\hat{\theta}(\mathbf{y}) - \mathbf{a}^T\theta\right)^2\right]$ is lower bounded by $\mathbf{a}^T F^{-1}(\theta)\mathbf{a}$, where $F(\theta)$ is the Fisher Information Matrix. The $(m,n)^{\text{th}}$ element of $F(\theta)$ is given by

$$F_{m,n}(\theta) = \mathbb{E}_{\mathbf{y}|\theta}\left\{\frac{\partial \ln p(\mathbf{y}|\theta)}{\partial \theta_m}\frac{\partial \ln p(\mathbf{y}|\theta)}{\partial \theta_n}\right\}. \tag{B.1}$$

For parameter estimation in additive white Gaussian noise i.e. $\mathbf{y} = \mathbf{s}(\theta) + \mathbf{z},\ \mathbf{z} \sim \mathcal{CN}(0, \sigma^2\mathbb{I})$, equation (B.1) simplifies to,

$$F_{m,n}(\theta) = \frac{2}{\sigma^2}\Re\left\{\left(\frac{\partial \mathbf{s}(\theta)}{\partial \theta_m}\right)^H \frac{\partial \mathbf{s}(\theta)}{\partial \theta_n}\right\}. \tag{B.2}$$

For our frequency estimation problem in (2.2), θ is the vector of all parameters, namely $\{|g_l|, \angle g_l, \omega_l : l = 1, \ldots, K\}$, and $\mathbf{s}(\theta) = \sum_{l=1}^{K} |g_l| e^{j\angle g_l}\mathbf{x}(\omega_l)$. We form $F(\theta)$ for this measurement model then choose those diagonal elements of $F^{-1}(\theta)$ that correspond to the frequencies $\{\omega_l : l = 1, \ldots, K\}$.

Appendix C

Proof of Theorem 4

The following lemmas will be used in the proof.

Lemma 2 *Let $\mathcal{T}_1$ and $\mathcal{T}_2$ be two sets of points in 2D space, with $\mathcal{T}_1 \subseteq \mathcal{T}_2$. Then, $\mathcal{I}_l(\mathcal{T}_1) \subseteq \mathcal{I}_l(\mathcal{T}_2)$.*

Proof: *For any $p_1 \in \mathcal{I}_l(\mathcal{T}_1)$, $\exists\ p_2 \in \mathcal{T}_1$, such that $p_1 = \mathcal{I}_l(p_2)$. Since $\mathcal{T}_1 \subseteq \mathcal{T}_2$, we have $p_2 \in \mathcal{T}_2$. Therefore, $p_1 \in \mathcal{I}_l(\mathcal{T}_2)$. This completes the proof.* □

Lemma 3 *The intersection of a circular segment with a line in 2D space, is either the empty set, or it contains at least one point from the arc boundary of the circular segment. The proof is simple and shown by Figure C.1.*

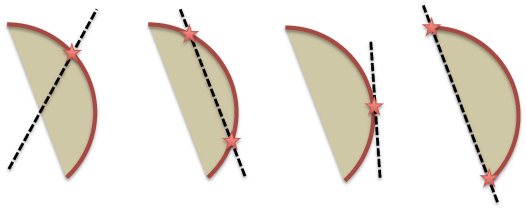


Figure C.1: Proof of Lemma 3. The intersection of the dashed line with the circular segment has at least one point on the arc, depicted by the star symbols.

Proof of Theorem 4: From (4.20) we have $\mathcal{T}_{\text{mono}} \subseteq \mathcal{T}_{\text{multi}} \subseteq \text{conv}(\mathcal{T}_{\text{mono}})$. By Lemma 2,

$$\mathcal{I}_l(\mathcal{T}_{\text{mono}}) \subseteq \mathcal{I}_l(\mathcal{T}_{\text{multi}}) \subseteq \mathcal{I}_l(\text{conv}(\mathcal{T}_{\text{mono}})). \tag{C.1}$$

Note that $\mathcal{T}_{\text{mono}}$ is an arc of a circle of radius $2k$, and its convex hull forms the corresponding circular segment. For any point $p_1 \in \mathcal{I}_l(\text{conv}(\mathcal{T}_{\text{mono}}))$, let us define the inverse image as $\mathcal{I}_l^{-1}(p_1) \triangleq \{p_2 \in \text{conv}(\mathcal{T}_{\text{mono}}) : \mathcal{I}_l(p_2) = p_1\}$. It is easy to see that $\mathcal{I}_l^{-1}(p_1)$ is the intersection of $\text{conv}(\mathcal{T}_{\text{mono}})$ with a line passing through p_1 and perpendicular to l. By lemma 3, $\mathcal{I}_l^{-1}(p_1)$ includes at least one point $p_2 \in \mathcal{T}_{\text{mono}}$, i.e., the arc boundary of $\text{conv}(\mathcal{T}_{\text{mono}})$. Therefore, $p_1 = \mathcal{I}_l(p_2) \in \mathcal{I}_l(\mathcal{T}_{\text{mono}})$, and hence, $\mathcal{I}_l(\text{conv}(\mathcal{T}_{\text{mono}})) \subseteq \mathcal{I}_l(\mathcal{T}_{\text{mono}})$. Combining this with $(C.1)$, we obtain

$$\mathcal{I}_l(\mathcal{T}_{\text{mono}}) = \mathcal{I}_l(\mathcal{T}_{\text{multi}}) = \mathcal{I}_l(\text{conv}(\mathcal{T}_{\text{mono}})). \tag{C.2}$$

This completes the proof. □

Appendix D

Effective Aperture Concept

Effective aperture (also known as virtual array) is a widely used technique for the design and analysis of multistatic arrays in far field [117, 118]. Based on this approach, one can derive an equivalent monostatic array for any given multistatic architecture by convolving the transmit and receive aperture functions [127] (defined in (4.39)). The classical analysis of the effective aperture relies on the notion of *array factor* (radiation pattern of the array in far field) defined by the Fourier transform of the aperture functions,

$$\begin{aligned} P_{tx}(g) &= \int_{\mathbb{R}} a_{tx}(x) e^{-j2\pi g x} dx, \\ P_{rx}(g) &= \int_{\mathbb{R}} a_{rx}(x) e^{-j2\pi g x} dx, \end{aligned} \tag{D.1}$$

where $g \triangleq \sin(\theta)$, θ being the angle measured from the perpendicular to the array. The two way array factor is given by the product of the transmit radiation pattern $P_{tx}(\cdot)$, and receive radiation pattern $P_{rx}(\cdot)$,

$$P_{eff}(g) = P_{tx}(g).P_{rx}(g). \tag{D.2}$$

Equivalently, the effective aperture is given by the convolution of the Tx and Rx aperture functions, i.e., $a_{tx}(x) * a_{rx}(x)$. This analysis does not capture the *shrinking* operation described in Subsection 4.5.2. Here, we show that the shrinkage and convolution operations in equation (4.41), is indeed consistent with the effective monostatic array (4.40) derived from Fresnel approximation;

$$\begin{aligned}
&a_{tx}(2x) * a_{rx}(2x) \\
&= \sum_{i=1}^{N_{tx}} \delta\left(x - \frac{x_{tx}(i)}{2}\right) * \sum_{j=1}^{N_{rx}} \delta\left(x - \frac{x_{rx}(j)}{2}\right) \\
&= \int_p \sum_{i=1}^{N_{tx}} \delta\left(x - \frac{x_{tx}(i)}{2} - p\right) . \sum_{j=1}^{N_{rx}} \delta\left(p - \frac{x_{rx}(j)}{2}\right) dp \\
&\overset{(a)}{=} \sum_{i=1}^{N_{tx}} \sum_{j=1}^{N_{rx}} \delta\left(x - \left(\frac{x_{tx}(i) + x_{rx}(j)}{2}\right)\right) = a_{eff}(x),
\end{aligned} \tag{D.3}$$

where (a) follows from the sifting property of the delta function.

Bibliography

[1] *Report and order and further notice of proposed rulemaking, federal communications commission, document FCC-16-89* (July, 2016).

[2] M. Murad, J. Nickolaou, G. Raz, J. S. Colburn, and K. Geary, *Next generation short range radar (srr) for automotive applications*, in *Radar Conference (RADAR), 2012 IEEE*, pp. 0214–0219, IEEE, 2012.

[3] Y. Zhu, Y. Zhu, B. Y. Zhao, and H. Zheng, *Reusing 60ghz radios for mobile radar imaging*, in *Proceedings of the 21st Annual International Conference on Mobile Computing and Networking*, pp. 103–116, ACM, 2015.

[4] J. Lien, N. Gillian, M. E. Karagozler, P. Amihood, C. Schwesig, E. Olson, H. Raja, and I. Poupyrev, *Soli: Ubiquitous gesture sensing with millimeter wave radar*, *ACM Transactions on Graphics (TOG)* **35** (2016), no. 4 142.

[5] T. Wei and X. Zhang, *mtrack: High-precision passive tracking using millimeter wave radios*, in *Proceedings of the 21st Annual International Conference on Mobile Computing and Networking*, pp. 117–129, ACM, 2015.

[6] F. Garcia, P. Cerri, A. Broggi, A. de la Escalera, and J. M. Armingol, *Data fusion for overtaking vehicle detection based on radar and optical flow*, in *2012 IEEE Intelligent Vehicles Symposium*, pp. 494–499, June, 2012.

[7] T. Wei, A. Zhou, and X. Zhang, *Facilitating robust 60 ghz network deployment by sensing ambient reflectors.*, in *NSDI*, pp. 213–226, 2017.

[8] F. Adib, C.-Y. Hsu, H. Mao, D. Katabi, and F. Durand, *Capturing the human figure through a wall*, *ACM Transactions on Graphics (TOG)* **34** (2015), no. 6 219.

[9] S. Depatla, L. Buckland, and Y. Mostofi, *X-Ray Vision with Only WiFi Power Measurements Using Rytov Wave Models*, *IEEE Transactions on Vehicular Technology* **64** (April, 2015) 1376–1387.

[10] C. R. Karanam and Y. Mostofi, *3d through-wall imaging with unmanned aerial vehicles using wifi*, in *Proceedings of the 16th ACM/IEEE International Conference on Information Processing in Sensor Networks*, pp. 131–142, ACM, 2017.

[11] S. Ahmed, A. Schiessl, F. Gumbmann, M. Tiebout, S. Methfessel, and L. Schmidt, *Advanced microwave imaging*, *Microwave Magazine, IEEE* **13** (Sept, 2012) 26–43.

[12] "NOMP Software." `https://bitbucket.org/wcslspectralestimation/continuous-frequency-estimation/src/NOMP`, 2016.

[13] J. A. Tropp and S. J. Wright, *Computational methods for sparse solution of linear inverse problems*, *Proceedings of the IEEE* **98** (2010), no. 6 948–958.

[14] J. A. Tropp and A. C. Gilbert, *Signal recovery from random measurements via orthogonal matching pursuit*, *Information Theory, IEEE Transactions on* **53** (2007), no. 12 4655–4666.

[15] Y. Chi, L. L. Scharf, A. Pezeshki, and A. R. Calderbank, *Sensitivity to basis mismatch in compressed sensing*, *Signal Processing, IEEE Transactions on* **59** (2011), no. 5 2182–2195.

[16] L. Jacques and C. De Vleeschouwer, *A geometrical study of matching pursuit parametrization*, *Signal Processing, IEEE Transactions on* **56** (2008), no. 7 2835–2848.

[17] S. G. Mallat and Z. Zhang, *Matching pursuits with time-frequency dictionaries*, *Signal Processing, IEEE Transactions on* **41** (1993), no. 12 3397–3415.

[18] H. L. Van Trees, *Detection, estimation, and modulation theory*. John Wiley & Sons, 2004.

[19] B. N. Bhaskar, G. Tang, and B. Recht, *Atomic norm denoising with applications to line spectral estimation*, *Signal Processing, IEEE Transactions on* **61** (2013), no. 23 5987–5999.

[20] R. O. Schmidt, *Multiple emitter location and signal parameter estimation*, *Antennas and Propagation, IEEE Transactions on* **34** (1986), no. 3 276–280.

[21] A. Arbabian, S. Callender, S. Kang, M. Rangwala, and A. Niknejad, *A 94 ghz mm-wave-to-baseband pulsed-radar transceiver with applications in imaging and gesture recognition*, *Solid-State Circuits, IEEE Journal of* **48** (April, 2013) 1055–1071.

[22] M. Steinhauer, H.-O. Ruob, H. Irion, and W. Menzel, *Millimeter-wave-radar sensor based on a transceiver array for automotive applications*, *IEEE transactions on microwave theory and techniques* **56** (2008), no. 2 261–269.

[23] C. Li, Z. Peng, T.-Y. Huang, T. Fan, F.-K. Wang, T.-S. Horng, J.-M. Muñoz-Ferreras, R. Gómez-García, L. Ran, and J. Lin, *A review on recent progress of portable short-range noncontact microwave radar systems*, *IEEE Transactions on Microwave Theory and Techniques* (2017).

[24] G. T. Di Francia, *Degrees of freedom of an image*, *JOSA* **59** (1969), no. 7 799–804.

[25] R. Piestun and D. A. Miller, *Electromagnetic degrees of freedom of an optical system*, *JOSA A* **17** (2000), no. 5 892–902.

[26] F. K. Gruber and E. A. Marengo, *New aspects of electromagnetic information theory for wireless and antenna systems*, *IEEE Transactions on Antennas and Propagation* **56** (Nov, 2008) 3470–3484.

[27] E. Torkildson, U. Madhow, and M. Rodwell, *Indoor millimeter wave mimo: Feasibility and performance*, *IEEE Transactions on Wireless Communications* **10** (2011), no. 12 4150–4160.

[28] B. Mamandipoor, M. Fallahpour, G. Malysa, K. Noujeim, U. Madhow, and A. Arbabian, *Spatial-domain technique to overcome grating lobes in sparse monostatic mm-wave imaging systems*, in *Microwave Symposium (IMS), 2016 IEEE MTT-S International*, pp. 1–4, IEEE, 2016.

[29] G. T. Di Francia, *Resolving power and information*, *JOSA* **45** (1955), no. 7 497–501.

[30] J. K. L. Tsang and K.-H. Ding, *Scattering of Electromagnetic Waves, Theories and Applications*. John Wiley and Sons, 2000.

[31] F. Gori and G. Guattari, *Shannon number and degrees of freedom of an image*, *Optics Communications* **7** (1973), no. 2 163–165.

[32] B. Mamandipoor, D. Ramasamy, and U. Madhow, *Newtonized orthogonal matching pursuit: Frequency estimation over the continuum*, *IEEE Transactions on Signal Processing* **64** (2016), no. 19 5066–5081.

[33] R. Roy and T. Kailath, *Esprit-estimation of signal parameters via rotational invariance techniques*, *Acoustics, Speech and Signal Processing, IEEE Transactions on* **37** (1989), no. 7 984–995.

[34] H. Teutsch, *Modal Array Signal Processing: Principles and Applications of Acoustic Wavefield Decomposition*. Lecture Notes in Control and Information Sciences. Springer Berlin Heidelberg, 2007.

[35] K. Duda, L. B. Magalas, M. Majewski, and T. P. Zieliński, *Dft-based estimation of damped oscillation parameters in low-frequency mechanical spectroscopy*, *Instrumentation and Measurement, IEEE Transactions on* **60** (2011), no. 11 3608–3618.

[36] T. Zieliński and K. Duda, *Frequency and damping estimation methods-an overview*, *Metrology and Measurement Systems* **18** (2011), no. 4 505–528.

[37] D. Vu, L. Xu, M. Xue, and J. Li, *Nonparametric missing sample spectral analysis and its applications to interrupted sar*, *Selected Topics in Signal Processing, IEEE Journal of* **6** (2012), no. 1 1–14.

[38] E. J. Candès and C. Fernandez-Granda, *Towards a mathematical theory of super-resolution*, *Communications on Pure and Applied Mathematics* **67** (2014), no. 6 906–956.

[39] C. Fernandez-Granda, *Super-resolution of point sources via convex programming*, *arXiv preprint arXiv:1507.07034* (2015).

[40] E. J. Candès and C. Fernandez-Granda, *Super-resolution from noisy data*, *Journal of Fourier Analysis and Applications* **19** (2013), no. 6 1229–1254.

[41] G. Tang, B. N. Bhaskar, and B. Recht, *Sparse recovery over continuous dictionaries-just discretize*, in *Signals, Systems and Computers, 2013 Asilomar Conference on*, pp. 1043–1047, IEEE, 2013.

[42] J. Friedman, T. Hastie, H. Höfling, and R. Tibshirani, *Pathwise coordinate optimization*, *The Annals of Applied Statistics* **1** (2007), no. 2 302–332.

[43] T. T. Wu and K. Lange, *Coordinate descent algorithms for lasso penalized regression*, *The Annals of Applied Statistics* (2008) 224–244.

[44] A. Panahi, M. Viberg, and B. Hassibi, *A numerical implementation of gridless compressed sensing*, in *Acoustics, Speech and Signal Processing (ICASSP), 2015 IEEE International Conference on*, pp. 3342–3346, IEEE, 2015.

[45] D. Needell and J. A. Tropp, *Cosamp: Iterative signal recovery from incomplete and inaccurate samples*, *Applied and Computational Harmonic Analysis* **26** (2009), no. 3 301–321.

[46] T. Zhang, *Adaptive forward-backward greedy algorithm for learning sparse representations*, *Information Theory, IEEE Transactions on* **57** (2011), no. 7 4689–4708.

[47] T. Blumensath and M. E. Davies, *Iterative thresholding for sparse approximations*, *Journal of Fourier Analysis and Applications* **14** (2008), no. 5-6 629–654.

[48] P. Jain, A. Tewari, and I. S. Dhillon, *Orthogonal matching pursuit with replacement*, in *Advances in Neural Information Processing Systems*, pp. 1215–1223, 2011.

[49] J. Li and P. Stoica, *Efficient mixed-spectrum estimation with applications to target feature extraction*, *Signal Processing, IEEE Transactions on* **44** (1996), no. 2 281–295.

[50] T. Yardibi, J. Li, P. Stoica, M. Xue, and A. B. Baggeroer, *Source localization and sensing: A nonparametric iterative adaptive approach based on weighted least squares*, Institute of Electrical and Electronics Engineers, 2010.

[51] A. Fannjiang and W. Liao, *Coherence pattern-guided compressive sensing with unresolved grids*, *SIAM Journal on Imaging Sciences* **5** (2012), no. 1 179–202.

[52] J. A. Tropp, *Greed is good: Algorithmic results for sparse approximation*, *Information Theory, IEEE Transactions on* **50** (2004), no. 10 2231–2242.

[53] M. A. Davenport and M. B. Wakin, *Analysis of orthogonal matching pursuit using the restricted isometry property*, *Information Theory, IEEE Transactions on* **56** (2010), no. 9 4395–4401.

[54] T. J. Abatzoglou, *A fast maximum likelihood algorithm for frequency estimation of a sinusoid based on newton's method*, *Acoustics, Speech and Signal Processing, IEEE Transactions on* **33** (1985), no. 1 77–89.

[55] P. Bidigare, U. Madhow, R. Mudumbai, and D. Scherber, *Attaining fundamental bounds on timing synchronization*, in *Acoustics, Speech and Signal Processing (ICASSP), 2012 IEEE International Conference on*, pp. 5229–5232, IEEE, 2012.

[56] D. Ramasamy, S. Venkateswaran, and U. Madhow, *Compressive adaptation of large steerable arrays*, in *Information Theory and Applications Workshop (ITA), 2012*, pp. 234–239, IEEE, 2012.

[57] D. Ramasamy, S. Venkateswaran, and U. Madhow, *Compressive tracking with 1000-element arrays: A framework for multi-gbps mm wave cellular downlinks*, in *Communication, Control, and Computing (Allerton), 2012 50th Annual Allerton Conference on*, pp. 690–697, IEEE, 2012.

[58] B. Mamandipoor, D. Ramasamy, and U. Madhow, *Frequency estimation for a mixture of sinusoids: A near-optimal sequential approach*, in *3rd IEEE Global Conference on Signal and Information Processing (GlobalSIP)*, 2015.

[59] T. L. Hansen, M. A. Badiu, B. H. Fleury, and B. Rao, *A sparse bayesian learning algorithm with dictionary parameter estimation*, in *Sensor Array and Multichannel Signal Processing Workshop (SAM), 2014 IEEE 8th*, pp. 385–388, IEEE, 2014.

[60] A. R. Barron, A. Cohen, W. Dahmen, and R. A. DeVore, *Approximation and learning by greedy algorithms*, *The annals of statistics* (2008) 64–94.

[61] D. Slepian, *Prolate spheroidal wave functions, fourier analysis, and uncertaintyv: The discrete case*, *Bell System Technical Journal* **57** (1978), no. 5 1371–1430.

[62] F. B. Hildebrand, *Introduction to numerical analysis*. Courier Corporation, 1987.

[63] H. L. Van Trees and K. L. Bell, *Bayesian bounds for parameter estimation and nonlinear filtering/tracking*, *AMC* **10** (2007) 12.

[64] D. Ramasamy, S. Venkateswaran, and U. Madhow, *Compressive parameter estimation in awgn*, *Signal Processing, IEEE Transactions on* **62** (2014), no. 8 2012–2027.

[65] P. P. Gandhi and S. A. Kassam, *Analysis of cfar processors in homogeneous background*, *Aerospace and Electronic Systems, IEEE Transactions on* **24** (1988), no. 4 427–445.

[66] H. Rohling, *Radar cfar thresholding in clutter and multiple target situations*, *Aerospace and Electronic Systems, IEEE Transactions on* (1983), no. 4 608–621.

[67] B. Eisenberg, *On the expectation of the maximum of iid geometric random variables*, *Statistics & Probability Letters* **78** (2008), no. 2 135–143.

[68] G. Schwarz, *Estimating the dimension of a model*, *The annals of statistics* **6** (1978), no. 2 461–464.

[69] H. Akaike, *A new look at the statistical model identification*, *Automatic Control, IEEE Transactions on* **19** (1974), no. 6 716–723.

[70] V. S. Ryaben'kii and S. V. Tsynkov, *A theoretical introduction to numerical analysis*. CRC Press, 2006.

[71] V. Chandrasekaran, B. Recht, P. A. Parrilo, and A. S. Willsky, *The convex geometry of linear inverse problems*, *Foundations of Computational mathematics* **12** (2012), no. 6 805–849.

[72] F. F. Bonsall, *A general atomic decomposition theorem and banach's closed range theorem*, *The Quarterly Journal of Mathematics* **42** (1991), no. 1 9–14.

[73] P. Shah, B. N. Bhaskar, G. Tang, and B. Recht, *Linear system identification via atomic norm regularization*, in *Decision and Control (CDC), 2012 IEEE 51st Annual Conference on*, pp. 6265–6270, 2012.

[74] M. Wax and T. Kailath, *Detection of signals by information theoretic criteria*, *Acoustics, Speech and Signal Processing, IEEE Transactions on* **33** (1985), no. 2 387–392.

[75] G. Tang, B. N. Bhaskar, and B. Recht, *Near minimax line spectral estimation*, *Information Theory, IEEE Transactions on* **61** (2015), no. 1 499–512.

[76] S. Boyd, N. Parikh, E. Chu, B. Peleato, and J. Eckstein, *Distributed optimization and statistical learning via the alternating direction method of multipliers*, *Foundations and Trends® in Machine Learning* **3** (2011), no. 1 1–122.

[77] S. J. Wright, R. D. Nowak, and M. A. Figueiredo, *Sparse reconstruction by separable approximation*, *Signal Processing, IEEE Transactions on* **57** (2009), no. 7 2479–2493.

[78] D. W. Hawkins, *Synthetic aperture imaging algorithms: with application to wide bandwidth sonar*. PhD thesis, University of Canterbury, Christchurch, New Zealand, 1996.

[79] J. Camacho, M. Parrilla, and C. Fritsch, *Phase coherence imaging.*, *IEEE transactions on ultrasonics, ferroelectrics, and frequency control* **56** (2009), no. 5 958–974.

[80] C. Martín, O. Martinez, L. Ullate, A. Octavio, and G. Godoy, *Reduction of grating lobes in saft images*, in *Ultrasonics Symposium, 2008. IUS 2008. IEEE*, pp. 721–724, IEEE, 2008.

[81] M. P. Hayes and P. T. Gough, *Broad-band synthetic aperture sonar*, *IEEE Journal of Oceanic Engineering* **17** (1992), no. 1 80–94.

[82] E. J. Luminati, *Wide-angle multistatic synthetic aperture radar: Focused image formation and aliasing artifact mitigation*. PhD thesis, Air Force Institute of Technology, Wright-Patterson Air Force Base, Ohio, USA, 2005.

[83] B. Mamandipoor, G. Malysa, A. Arbabian, U. Madhow, and K. Noujeim, *60 ghz synthetic aperture radar for short-range imaging: Theory and experiments*, in *Signals, Systems and Computers, 2014 48th Asilomar Conference on*, pp. 553–558, IEEE, 2014.

[84] M. Cetin, I. Stojanovic, N. Onhon, K. Varshney, S. Samadi, W. Karl, and A. Willsky, *Sparsity-driven synthetic aperture radar imaging: Reconstruction, autofocusing, moving targets, and compressed sensing*, *Signal Processing Magazine, IEEE* **31** (July, 2014) 27–40.

[85] S. Samadi, M. Cetin, and M. Masnadi-Shirazi, *Sparse representation-based synthetic aperture radar imaging*, *Radar, Sonar Navigation, IET* **5** (Feb, 2011) 182–193.

[86] S. Samadi, M. Cetin, and M. A. Masnadi-Shirazi, *Multiple feature-enhanced sar imaging using sparsity in combined dictionaries*, *IEEE Geoscience and Remote Sensing Letters* **10** (July, 2013) 821–825.

[87] S. S. Chen, D. L. Donoho, and M. A. Saunders, *Atomic decomposition by basis pursuit*, *SIAM review* **43** (2001), no. 1 129–159.

[88] S. Foucart and M.-J. Lai, *Sparsest solutions of underdetermined linear systems via ℓ_q-minimization for $0 < q \leq 1$*, *Applied and Computational Harmonic Analysis* **26** (2009), no. 3 395–407.

[89] R. Tibshirani, *Regression shrinkage and selection via the lasso*, *Journal of the Royal Statistical Society. Series B (Methodological)* (1996) 267–288.

[90] W. Southwell, *Validity of the fresnel approximation in the near field*, *JOSA* **71** (1981), no. 1 7–14.

[91] A. Thaning, P. Martinsson, M. Karelin, and A. T. Friberg, *Limits of diffractive optics by communication modes*, *Journal of Optics A: Pure and Applied Optics* **5** (2003), no. 3 153.

[92] B. Mamandipoor, A. Arbabian, and U. Madhow, *New models and super-resolution techniques for short-range radar: Theory and experiments*, in *2016 Information Theory and Applications Workshop (ITA)*, pp. 1–7, Jan, 2016.

[93] K. Noujeim, G. Malysa, A. Babveyh, and A. Arbabian, *A compact nonlinear-transmission-line-based mm-wave sfcw imaging radar*, in *44th European Microwave Conference (EuMC)*, pp. 1766–1769, Oct, 2014.

[94] R. Bamler, *Principles of synthetic aperture radar*, *Surveys in Geophysics* **21** (2000), no. 2-3 147–157.

[95] S. S. Ahmed, A. Schiessl, and L.-P. Schmidt, *Illumination properties of multistatic planar arrays in near-field imaging applications*, in *Radar Conference (EuRAD), 2010 European*, pp. 29–32, IEEE, 2010.

[96] R. Baraniuk and P. Steeghs, *Compressive radar imaging*, in *2007 IEEE Radar Conference*, pp. 128–133, April, 2007.

[97] E. J. Candès, J. Romberg, and T. Tao, *Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information*, *IEEE Transactions on information theory* **52** (2006), no. 2 489–509.

[98] G. Di Francia, *Directivity, super-gain and information*, *IRE Transactions on Antennas and Propagation* **4** (1956), no. 3 473–478.

[99] H. Wolter, *V on basic analogies and principal differences between optical and electronic information*, *Progress in Optics* **1** (1961) 155–210.

[100] D. Slepian, *On bandwidth*, *Proceedings of the IEEE* **64** (1976), no. 3 292–300.

[101] H. J. Landau and H. O. Pollak, *Prolate spheroidal wave functions, fourier analysis and uncertainty – iii: The dimension of the space of essentially time-and band-limited signals*, *Bell Labs Technical Journal* **41** (1962), no. 4 1295–1336.

[102] H. Lee, *Acoustical Sensing and Imaging*. CRC Press, 2016.

[103] R. Pierri and F. Soldovieri, *On the information content of the radiated fields in the near zone over bounded domains*, *Inverse Problems* **14** (1998), no. 2 321.

[104] R. Solimene, G. Leone, and R. Pierri, *Resolution in two-dimensional tomographic reconstructions in the fresnel zone from born scattered fields*, *Journal of Optics A: Pure and Applied Optics* **6** (2004), no. 6 529.

[105] R. Solimene, C. Mola, G. Gennarelli, and F. Soldovieri, *On the singular spectrum of radiation operators in the non-reactive zone: the case of strip sources*, *Journal of Optics* **17** (2015), no. 2 025605.

[106] R. Solimene, M. A. Maisto, G. Romeo, and R. Pierri, *On the singular spectrum of the radiation operator for multiple and extended observation domains*, *International Journal of Antennas and Propagation* **2013** (2013).

[107] R. Pierri, A. Liseno, R. Solimene, and F. Tartaglione, *In-depth resolution from multifrequency born fields scattered by a dielectric strip in the fresnel zone*, *JOSA A* **19** (2002), no. 6 1234–1238.

[108] A. W. Lohmann, R. G. Dorsch, D. Mendlovic, C. Ferreira, and Z. Zalevsky, *Space–bandwidth product of optical signals and systems*, *JOSA A* **13** (1996), no. 3 470–473.

[109] D. Mendlovic and A. W. Lohmann, *Space–bandwidth product adaptation and its application to superresolution: fundamentals*, *JOSA A* **14** (1997), no. 3 558–562.

[110] M. Bastiaans, *Wigner distribution function and its application to first-order optics*, *JOSA* **69** (1979), no. 12 1710–1716.

[111] D. K. Cheng *et. al.*, *Field and wave electromagnetics*. Pearson Education India, 1989.

[112] M. Bertero and P. Boccacci, *Introduction to inverse problems in imaging*. CRC press, 1998.

[113] D. Porter and D. S. Stirling, *Integral equations: a practical treatment, from spectral theory to applications*, vol. 5. Cambridge University Press, 1990.

[114] W. C. Chew, *Waves and fields in inhomogeneous media*, vol. 522. IEEE press New York, 1995.

[115] S. S. Ahmed, *Electronic microwave imaging with planar multistatic arrays*. Logos Verlag Berlin GmbH, 2014.

[116] R. Solimene, G. Leone, and R. Pierri, *Multistatic–multiview resolution from born fields for strips in fresnel zone*, *JOSA A* **21** (2004), no. 8 1402–1406.

[117] G. Lockwood and F. S. Foster, *Optimizing sparse two-dimensional transducer arrays using an effective aperture approach*, in *Ultrasonics Symposium, 1994. Proceedings., 1994 IEEE*, vol. 3, pp. 1497–1501, IEEE, 1994.

[118] S. S. Ahmed, A. Schiess, and L.-P. Schmidt, *Near field mm-wave imaging with multistatic sparse 2d-arrays*, in *Radar Conference, 2009. EuRAD 2009. European*, pp. 180–183, IEEE, 2009.

[119] A. Louis, *Medical imaging: state of the art and future development*, *Inverse Problems* **8** (1992), no. 5 709.

[120] T. S. Ralston, D. L. Marks, P. S. Carney, and S. A. Boppart, *Inverse scattering for optical coherence tomography*, *JOSA A* **23** (2006), no. 5 1027–1037.

[121] F. Cakoni and D. Colton, *Qualitative methods in inverse scattering theory: An introduction*. Springer Science & Business Media, 2005.

[122] C. J. Nolan and M. Cheney, *Synthetic aperture inversion*, *Inverse Problems* **18** (2002), no. 1 221.

[123] S. S. Ahmed, A. Schiessl, and L.-P. Schmidt, *Multistatic mm-wave imaging with planar 2d-arrays*, in *Microwave Conference, 2009 German*, pp. 1–4, IEEE, 2009.

[124] A. Den Dekker and A. Van den Bos, *Resolution: a survey*, *JOSA A* **14** (1997), no. 3 547–557.

[125] Z. Zalevsky and D. Mendlovic, *Optical superresolution*, vol. 91. Springer, 2004.

[126] A. Gupta, U. Madhow, and A. Arbabian, *Super-resolution in position and velocity estimation for short-range mm-wave radar*, in *2016 50th Asilomar Conference on Signals, Systems and Computers*, pp. 1144–1148, Nov, 2016.

[127] P. Pal and P. Vaidyanathan, *Nested arrays: A novel approach to array processing with enhanced degrees of freedom*, *IEEE Transactions on Signal Processing* **58** (2010), no. 8 4167–4181.